

1. [Preface](#)
2. Introduction: The Nature of Science and Physics
 1. [Introduction to Science and the Realm of Physics, Physical Quantities, and Units](#)
 2. [Physics: An Introduction](#)
 3. [Physical Quantities and Units](#)
 4. [Accuracy, Precision, and Significant Figures](#)
 5. [Approximation](#)
3. Kinematics
 1. [Introduction to One-Dimensional Kinematics](#)
 2. [Displacement](#)
 3. [Vectors, Scalars, and Coordinate Systems](#)
 4. [Time, Velocity, and Speed](#)
 5. [Acceleration](#)
 6. [Motion Equations for Constant Acceleration in One Dimension](#)
 7. [Problem-Solving Basics for One-Dimensional Kinematics](#)
 8. [Falling Objects](#)
 9. [Graphical Analysis of One-Dimensional Motion](#)
4. Dynamics: Force and Newton's Laws of Motion
 1. [Introduction to Dynamics: Newton's Laws of Motion](#)
 2. [Development of Force Concept](#)
 3. [Newton's First Law of Motion: Inertia](#)
 4. [Newton's Second Law of Motion: Concept of a System](#)
 5. [Newton's Third Law of Motion: Symmetry in Forces](#)
 6. [Normal, Tension, and Other Examples of Forces](#)
 7. [Problem-Solving Strategies](#)
 8. [Further Applications of Newton's Laws of Motion](#)
 9. [Extended Topic: The Four Basic Forces—An Introduction](#)
5. Further Applications of Newton's Laws: Friction, Drag, and Elasticity

1. [Introduction: Further Applications of Newton's Laws](#)
 2. [Friction](#)
 3. [Drag Forces](#)
6. Uniform Circular Motion and Gravitation
 1. [Newton's Universal Law of Gravitation](#)
 2. [Satellites and Kepler's Laws: An Argument for Simplicity](#)
7. Work, Energy, and Energy Resources
 1. [Introduction to Work, Energy, and Energy Resources](#)
 2. [Work: The Scientific Definition](#)
 3. [Kinetic Energy and the Work-Energy Theorem](#)
 4. [Gravitational Potential Energy](#)
 5. [Conservative Forces and Potential Energy](#)
 6. [Nonconservative Forces](#)
 7. [Conservation of Energy](#)
 8. [Power](#)
 9. [Work, Energy, and Power in Humans](#)
 10. [World Energy Use](#)
8. Linear Momentum and Collisions
 1. [Introduction to Linear Momentum and Collisions](#)
 2. [Linear Momentum and Force](#)
 3. [Impulse](#)
 4. [Conservation of Momentum](#)
 5. [Elastic Collisions in One Dimension](#)
 6. [Inelastic Collisions in One Dimension](#)
9. Statics and Torque
 1. [Introduction to Statics and Torque](#)
 2. [The First Condition for Equilibrium](#)
 3. [The Second Condition for Equilibrium](#)
 4. [Applications of Statics, Including Problem-Solving Strategies](#)
 5. [Simple Machines](#)
10. Temperature, Kinetic Theory, and the Gas Laws

1. [Introduction to Temperature, Kinetic Theory, and the Gas Laws](#)
2. [Temperature](#)
3. [Thermal Expansion of Solids and Liquids](#)
4. [The Ideal Gas Law](#)
5. [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#)
6. [Phase Changes](#)
7. [Humidity, Evaporation, and Boiling](#)

11. Heat and Heat Transfer Methods

1. [Introduction to Heat and Heat Transfer Methods](#)
2. [Heat](#)
3. [Temperature Change and Heat Capacity](#)
4. [Phase Change and Latent Heat](#)
5. [Heat Transfer Methods](#)
6. [Conduction](#)
7. [Convection](#)
8. [Radiation](#)

12. Thermodynamics

1. [Introduction to Thermodynamics](#)
2. [The First Law of Thermodynamics](#)
3. [The First Law of Thermodynamics and Some Simple Processes](#)
4. [Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency](#)
5. [Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated](#)
6. [Applications of Thermodynamics: Heat Pumps and Refrigerators](#)
7. [Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy](#)

8. [Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation](#)

13. Oscillatory Motion and Waves

1. [Introduction to Oscillatory Motion and Waves](#)
2. [Hooke's Law: Stress and Strain Revisited](#)
3. [Period and Frequency in Oscillations](#)
4. [Simple Harmonic Motion: A Special Periodic Motion](#)
5. [The Simple Pendulum](#)
6. [Energy and the Simple Harmonic Oscillator](#)
7. [Forced Oscillations and Resonance](#)
8. [Waves](#)
9. [Superposition and Interference](#)
10. [Energy in Waves: Intensity](#)

14. Electric Charge and Electric Field

1. [Introduction to Electric Charge and Electric Field](#)
2. [Static Electricity and Charge: Conservation of Charge](#)
3. [Conductors and Insulators](#)
4. [Coulomb's Law](#)
5. [Electric Field: Concept of a Field Revisited](#)
6. [Electric Field Lines: Multiple Charges](#)
7. [Electric Forces in Biology](#)
8. [Conductors and Electric Fields in Static Equilibrium](#)
9. [Applications of Electrostatics](#)

15. Electric Potential and Electric Field

1. [Introduction to Electric Potential and Electric Energy](#)
2. [Electric Potential Energy: Potential Difference](#)
3. [Electric Potential in a Uniform Electric Field](#)
4. [Electrical Potential Due to a Point Charge](#)
5. [Equipotential Lines](#)

16. Electric Current, Resistance, and Ohm's Law

1. [Introduction to Electric Current, Resistance, and Ohm's Law](#)

2. [Current](#)
3. [Ohm's Law: Resistance and Simple Circuits](#)
4. [Resistance and Resistivity](#)
5. [Electric Power and Energy](#)
6. [Alternating Current versus Direct Current](#)
7. [Electric Hazards and the Human Body](#)
8. [Nerve Conduction–Electrocardiograms](#)
17. Circuits, Bioelectricity, and DC Instruments
 1. [Introduction to Circuits and DC Instruments](#)
 2. [Resistors in Series and Parallel](#)
 3. [Electromotive Force: Terminal Voltage](#)
 4. [Kirchhoff's Rules](#)
 5. [DC Voltmeters and Ammeters](#)
 6. [Null Measurements](#)
18. Magnetism
 1. [Introduction to Magnetism](#)
 2. [Magnets](#)
 3. [Ferromagnets and Electromagnets](#)
 4. [Magnetic Fields and Magnetic Field Lines](#)
 5. [Magnetic Field Strength: Force on a Moving Charge in a Magnetic Field](#)
 6. [Force on a Moving Charge in a Magnetic Field: Examples and Applications](#)
 7. [The Hall Effect](#)
 8. [Magnetic Force on a Current-Carrying Conductor](#)
 9. [Torque on a Current Loop: Motors and Meters](#)
 10. [Magnetic Fields Produced by Currents: Ampere's Law](#)
 11. [Magnetic Force between Two Parallel Conductors](#)
 12. [More Applications of Magnetism](#)
19. Electromagnetic Waves
 1. [Introduction to Electromagnetic Waves](#)

2. [Maxwell's Equations: Electromagnetic Waves Predicted and Observed](#)
3. [Production of Electromagnetic Waves](#)
4. [The Electromagnetic Spectrum](#)
5. [Energy in Electromagnetic Waves](#)

20. Geometric Optics

1. [Introduction to Geometric Optics](#)
2. [The Ray Aspect of Light](#)
3. [The Law of Reflection](#)
4. [The Law of Refraction](#)
5. [Total Internal Reflection](#)
6. [Dispersion: The Rainbow and Prisms](#)
7. [Image Formation by Lenses](#)
8. [Image Formation by Mirrors](#)

21. Vision and Optical Instruments

1. [Introduction to Vision and Optical Instruments](#)
2. [Physics of the Eye](#)
3. [Vision Correction](#)
4. [Color and Color Vision](#)
5. [Microscopes](#)
6. [Telescopes](#)
7. [Aberrations](#)

22. Introduction to Quantum Physics

1. [Introduction to Quantum Physics](#)
2. [Quantization of Energy](#)
3. [The Photoelectric Effect](#)
4. [Photon Energies and the Electromagnetic Spectrum](#)
5. [Photon Momentum](#)
6. [The Particle-Wave Duality](#)
7. [The Wave Nature of Matter](#)
8. [Probability: The Heisenberg Uncertainty Principle](#)
9. [The Particle-Wave Duality Reviewed](#)

23. Frontiers of Physics

1. [Introduction to Frontiers of Physics](#)
 2. [Cosmology and Particle Physics](#)
 3. [General Relativity and Quantum Gravity](#)
 4. [Superstrings](#)
 5. [Dark Matter and Closure](#)
 6. [Complexity and Chaos](#)
 7. [High-temperature Superconductors](#)
 8. [Some Questions We Know to Ask](#)
24. [Atomic Masses](#)
 25. [Selected Radioactive Isotopes](#)
 26. [Useful Information](#)
 27. [Glossary of Key Symbols and Notation](#)

Preface

Welcome to *College Physics*, an OpenStax resource. This textbook was written to increase student access to high-quality learning materials, maintaining highest standards of academic rigor at little to no cost.

About OpenStax

OpenStax is a nonprofit based at Rice University, and it's our mission to improve student access to education. Our first openly licensed college textbook was published in 2012, and our library has since scaled to over 20 books for college and AP courses used by hundreds of thousands of students. Our adaptive learning technology, designed to improve learning outcomes through personalized educational paths, is being piloted in college courses throughout the country. Through our partnerships with philanthropic foundations and our alliance with other educational resource organizations, OpenStax is breaking down the most common barriers to learning and empowering students and instructors to succeed.

About OpenStax Resources

Customization

College Physics is licensed under a Creative Commons Attribution 4.0 International (CC BY) license, which means that you can distribute, remix, and build upon the content, as long as you provide attribution to OpenStax and its content contributors.

Because our books are openly licensed, you are free to use the entire book or pick and choose the sections that are most relevant to the needs of your course. Feel free to remix the content by assigning your students certain chapters and sections in your syllabus, in the order that you prefer. You can even provide a direct link in your syllabus to the sections in the web view of your book.

Instructors also have the option of creating a customized version of their OpenStax book. The custom version can be made available to students in low-cost print or digital form through their campus bookstore. Visit your book page on openstax.org for more information.

Errata

All OpenStax textbooks undergo a rigorous review process. However, like any professional-grade textbook, errors sometimes occur. Since our books are web based, we can make updates periodically when deemed pedagogically necessary. If you have a correction to suggest, submit it through the link on your book page on openstax.org. Subject matter experts review all errata suggestions. OpenStax is committed to remaining transparent about all updates, so you will also find a list of past errata changes on your book page on openstax.org.

Format

You can access this textbook for free in web view or PDF through openstax.org, and in low-cost print and iBooks editions.

About *College Physics*

College Physics meets standard scope and sequence requirements for a two-semester introductory algebra-based physics course. The text is grounded in real-world examples to help students grasp fundamental physics concepts. It requires knowledge of algebra and some trigonometry, but not calculus. *College Physics* includes learning objectives, concept questions, links to labs and simulations, and ample practice opportunities for traditional physics application problems.

Coverage and Scope

College Physics is organized such that topics are introduced conceptually with a steady progression to precise definitions and analytical applications. The analytical aspect (problem solving) is tied back to the conceptual before moving on to another topic. Each introductory chapter, for example, opens with an engaging photograph relevant to the subject of the chapter and interesting applications that are easy for most students to visualize.

Chapter 1: Introduction: The Nature of Science and Physics

Chapter 2: Kinematics

Chapter 3: Two-Dimensional Kinematics

Chapter 4: Dynamics: Force and Newton's Laws of Motion

Chapter 5: Further Applications of Newton's Laws: Friction, Drag, and Elasticity

Chapter 6: Uniform Circular Motion and Gravitation

Chapter 7: Work, Energy, and Energy Resources

Chapter 8: Linear Momentum and Collisions

Chapter 9: Statics and Torque

Chapter 10: Rotational Motion and Angular Momentum

Chapter 11: Fluid Statics

Chapter 12: Fluid Dynamics and Its Biological and Medical Applications

Chapter 13: Temperature, Kinetic Theory, and the Gas Laws

Chapter 14: Heat and Heat Transfer Methods

Chapter 15: Thermodynamics

Chapter 16: Oscillatory Motion and Waves

Chapter 17: Physics of Hearing

Chapter 18: Electric Charge and Electric Field

Chapter 19: Electric Potential and Electric Field

Chapter 20: Electric Current, Resistance, and Ohm's Law

Chapter 21: Circuits and DC Instruments

Chapter 22: Magnetism

Chapter 23: Electromagnetic Induction, AC Circuits, and Electrical Technologies

Chapter 24: Electromagnetic Waves

Chapter 25: Geometric Optics

Chapter 26: Vision and Optical Instruments

Chapter 27: Wave Optics

Chapter 28: Special Relativity

Chapter 29: Introduction to Quantum Physics
Chapter 30: Atomic Physics
Chapter 31: Radioactivity and Nuclear Physics
Chapter 32: Medical Applications of Nuclear Physics
Chapter 33: Particle Physics
Chapter 34: Frontiers of Physics
Appendix A: Atomic Masses
Appendix B: Selected Radioactive Isotopes
Appendix C: Useful Information
Appendix D: Glossary of Key Symbols and Notation

Concepts and Calculations

The ability to calculate does not guarantee conceptual understanding. In order to unify conceptual, analytical, and calculation skills within the learning process, we have integrated Strategies and Discussions throughout the text.

Modern Perspective

The chapters on modern physics are more complete than many other texts on the market, with an entire chapter devoted to medical applications of nuclear physics and another to particle physics. The final chapter of the text, “Frontiers of Physics,” is devoted to the most exciting endeavors in physics. It ends with a module titled “Some Questions We Know to Ask.”

Key Features

Modularity

This textbook is organized as a collection of modules that can be rearranged and modified to suit the needs of a particular professor or class. That being said, modules often contain references to content in other modules, as most topics in physics cannot be discussed in isolation.

Learning Objectives

Every module begins with a set of learning objectives. These objectives are designed to guide the instructor in deciding what content to include or assign, and to guide the student with respect to what he or she can expect to learn. After completing the module and end-of-module exercises, students should be able to demonstrate mastery of the learning objectives.

Call-Outs

Key definitions, concepts, and equations are called out with a special design treatment. Call-outs are designed to catch readers' attention, to make it clear that a specific term, concept, or equation is particularly important, and to provide easy reference for a student reviewing content.

Key Terms

Key terms are in bold and are followed by a definition in context. Definitions of key terms are also listed in the Glossary, which appears at the end of the module.

Worked Examples

Worked examples have four distinct parts to promote both analytical and conceptual skills. Worked examples are introduced in words, always using some application that should be of interest. This is followed by a Strategy section that emphasizes the concepts involved and how solving the problem

relates to those concepts. This is followed by the mathematical Solution and Discussion.

Many worked examples contain multiple-part problems to help the students learn how to approach normal situations, in which problems tend to have multiple parts. Finally, worked examples employ the techniques of the problem-solving strategies so that students can see how those strategies succeed in practice as well as in theory.

Problem-Solving Strategies

Problem-solving strategies are first presented in a special section and subsequently appear at crucial points in the text where students can benefit most from them. Problem-solving strategies have a logical structure that is reinforced in the worked examples and supported in certain places by line drawings that illustrate various steps.

Misconception Alerts

Students come to physics with preconceptions from everyday experiences and from previous courses. Some of these preconceptions are misconceptions, and many are very common among students and the general public. Some are inadvertently picked up through misunderstandings of lectures and texts. The Misconception Alerts feature is designed to point these out and correct them explicitly.

Take-Home Investigations

Take Home Investigations provide the opportunity for students to apply or explore what they have learned with a hands-on activity.

Things Great and Small

In these special topic essays, macroscopic phenomena (such as air pressure) are explained with submicroscopic phenomena (such as atoms bouncing off walls). These essays support the modern perspective by describing aspects of modern physics before they are formally treated in later chapters. Connections are also made between apparently disparate phenomena.

Simulations

Where applicable, students are directed to the interactive PHeT physics simulations developed by the University of Colorado. There they can further explore the physics concepts they have learned about in the module.

Summary

Module summaries are thorough and functional and present all important definitions and equations. Students are able to find the definitions of all terms and symbols as well as their physical relationships. The structure of the summary makes plain the fundamental principles of the module or collection and serves as a useful study guide.

Glossary

At the end of every module or chapter is a Glossary containing definitions of all of the key terms in the module or chapter.

End-of-Module Problems

At the end of every chapter is a set of Conceptual Questions and/or skills-based Problems & Exercises. Conceptual Questions challenge students' ability to explain what they have learned conceptually, independent of the mathematical details. Problems & Exercises challenge students to apply both concepts and skills to solve mathematical physics problems. Online,

every other problem includes an answer that students can reveal immediately by clicking on a “Show Solution” button.

In addition to traditional skills-based problems, there are three special types of end-of-module problems: Integrated Concept Problems, Unreasonable Results Problems, and Construct Your Own Problems. All of these problems are indicated with a subtitle preceding the problem.

Integrated Concept Problems

In Integrated Concept Problems, students are asked to apply what they have learned about two or more concepts to arrive at a solution to a problem. These problems require a higher level of thinking because, before solving a problem, students have to recognize the combination of strategies required to solve it.

Unreasonable Results

In Unreasonable Results Problems, students are challenged to not only apply concepts and skills to solve a problem, but also to analyze the answer with respect to how likely or realistic it really is. These problems contain a premise that produces an unreasonable answer and are designed to further emphasize that properly applied physics must describe nature accurately and is not simply the process of solving equations.

Construct Your Own Problem

These problems require students to construct the details of a problem, justify their starting assumptions, show specific steps in the problem’s solution, and finally discuss the meaning of the result. These types of problems relate well to both conceptual and analytical aspects of physics, emphasizing that physics must describe nature. Often they involve an integration of topics from more than one chapter. Unlike other problems, solutions are not provided since there is no single correct answer.

Instructors should feel free to direct students regarding the level and scope of their considerations. Whether the problem is solved and described correctly will depend on initial assumptions.

Additional Resources

Student and Instructor Resources

We've compiled additional resources for both students and instructors, including Getting Started Guides, an instructor solution manual, and PowerPoint slides. Instructor resources require a verified instructor account, which can be requested on your openstax.org log-in. Take advantage of these resources to supplement your OpenStax book.

Partner Resources

OpenStax Partners are our allies in the mission to make high-quality learning materials affordable and accessible to students and instructors everywhere. Their tools integrate seamlessly with our OpenStax titles at a low cost. To access the partner resources for your text, visit your book page on openstax.org.

About the Authors

Senior Contributing Authors

Paul Peter Urone, Professor Emeritus at California State University, Sacramento

Roger Hinrichs, State University of New York, College at Oswego

Contributing Authors

Kim Dirks, University of Auckland
Manjula Sharma, University of Sydney

Reviewers

Matthew Adams, Crafton Hills College, San Bernardino Community
College District
Erik Christensen, South Florida Community College
Douglas Ingram, Texas Christian University
Eric Kincanon, Gonzaga University
Lee H. LaRue, Paris Junior College
Chuck Pearson, Virginia Intermont College
Marc Sher, College of William and Mary
Ulrich Zurcher, Cleveland State University

Introduction to Science and the Realm of Physics, Physical Quantities, and Units

class="introduction"

Galaxies are
as immense
as atoms are
small. Yet the
same laws of
physics
describe
both, and all
the rest of
nature—an
indication of
the
underlying
unity in the
universe. The
laws of
physics are
surprisingly
few in
number,
implying an
underlying
simplicity to
nature's
apparent
complexity.
(credit:
NASA, JPL-
Caltech, P.
Barmby,
Harvard-
Smithsonian
Center for

Astrophysics)



What is your first reaction when you hear the word “physics”? Did you imagine working through difficult equations or memorizing formulas that seem to have no real use in life outside the physics classroom? Many people come to the subject of physics with a bit of fear. But as you begin your exploration of this broad-ranging subject, you may soon come to realize that physics plays a much larger role in your life than you first thought, no matter your life goals or career choice.

For example, take a look at the image above. This image is of the Andromeda Galaxy, which contains billions of individual stars, huge clouds of gas, and dust. Two smaller galaxies are also visible as bright blue spots in the background. At a staggering 2.5 million light years from the Earth, this galaxy is the nearest one to our own galaxy (which is called the Milky Way). The stars and planets that make up Andromeda might seem to be the furthest thing from most people’s regular, everyday lives. But Andromeda is a great starting point to think about the forces that hold together the universe. The forces that cause Andromeda to act as it does are the same forces we contend with here on Earth, whether we are planning to send a rocket into space or simply raise the walls for a new home. The same gravity that causes the stars of Andromeda to rotate and revolve also causes water to flow over hydroelectric dams here on Earth. Tonight, take a moment to look up at the stars. The forces out there are the same as the ones here on Earth. Through a study of physics, you may gain a greater

understanding of the interconnectedness of everything we can see and know in this universe.

Think now about all of the technological devices that you use on a regular basis. Computers, smart phones, GPS systems, MP3 players, and satellite radio might come to mind. Next, think about the most exciting modern technologies that you have heard about in the news, such as trains that levitate above tracks, “invisibility cloaks” that bend light around them, and microscopic robots that fight cancer cells in our bodies. All of these groundbreaking advancements, commonplace or unbelievable, rely on the principles of physics. Aside from playing a significant role in technology, professionals such as engineers, pilots, physicians, physical therapists, electricians, and computer programmers apply physics concepts in their daily work. For example, a pilot must understand how wind forces affect a flight path and a physical therapist must understand how the muscles in the body experience forces as they move and bend. As you will learn in this text, physics principles are propelling new, exciting technologies, and these principles are applied in a wide range of careers.

In this text, you will begin to explore the history of the formal study of physics, beginning with natural philosophy and the ancient Greeks, and leading up through a review of Sir Isaac Newton and the laws of physics that bear his name. You will also be introduced to the standards scientists use when they study physical quantities and the interrelated system of measurements most of the scientific community uses to communicate in a single mathematical language. Finally, you will study the limits of our ability to be accurate and precise, and the reasons scientists go to painstaking lengths to be as clear as possible regarding their own limitations.

Physics: An Introduction

- Explain the difference between a principle and a law.
- Explain the difference between a model and a theory.



The flight formations of migratory birds such as Canada geese are governed by the laws of physics.
(credit: David Merrett)

The physical universe is enormously complex in its detail. Every day, each of us observes a great variety of objects and phenomena. Over the centuries, the curiosity of the human race has led us collectively to explore and catalog a tremendous wealth of information. From the flight of birds to the colors of flowers, from lightning to gravity, from quarks to clusters of galaxies, from the flow of time to the mystery of the creation of the universe, we have asked questions and assembled huge arrays of facts. In the face of all these details, we have discovered that a surprisingly small and unified set of physical laws can explain what we observe. As humans, we make generalizations and seek order. We have found that nature is remarkably cooperative—it exhibits the *underlying order and simplicity* we so value.

It is the underlying order of nature that makes science in general, and physics in particular, so enjoyable to study. For example, what do a bag of chips and a car battery have in common? Both contain energy that can be

converted to other forms. The law of conservation of energy (which says that energy can change form but is never lost) ties together such topics as food calories, batteries, heat, light, and watch springs. Understanding this law makes it easier to learn about the various forms energy takes and how they relate to one another. Apparently unrelated topics are connected through broadly applicable physical laws, permitting an understanding beyond just the memorization of lists of facts.

The unifying aspect of physical laws and the basic simplicity of nature form the underlying themes of this text. In learning to apply these laws, you will, of course, study the most important topics in physics. More importantly, you will gain analytical abilities that will enable you to apply these laws far beyond the scope of what can be included in a single book. These analytical skills will help you to excel academically, and they will also help you to think critically in any professional career you choose to pursue. This module discusses the realm of physics (to define what physics is), some applications of physics (to illustrate its relevance to other disciplines), and more precisely what constitutes a physical law (to illuminate the importance of experimentation to theory).

Science and the Realm of Physics

Science consists of the theories and laws that are the general truths of nature as well as the body of knowledge they encompass. Scientists are continually trying to expand this body of knowledge and to perfect the expression of the laws that describe it. **Physics** is concerned with describing the interactions of energy, matter, space, and time, and it is especially interested in what fundamental mechanisms underlie every phenomenon. The concern for describing the basic phenomena in nature essentially defines the *realm of physics*.

Physics aims to describe the function of everything around us, from the movement of tiny charged particles to the motion of people, cars, and spaceships. In fact, almost everything around you can be described quite accurately by the laws of physics. Consider a smart phone ([\[link\]](#)). Physics describes how electricity interacts with the various circuits inside the device. This knowledge helps engineers select the appropriate materials and

circuit layout when building the smart phone. Next, consider a GPS system. Physics describes the relationship between the speed of an object, the distance over which it travels, and the time it takes to travel that distance. When you use a GPS device in a vehicle, it utilizes these physics equations to determine the travel time from one location to another.



The Apple
“iPhone” is a
common
smart phone
with a GPS
function.

Physics
describes the
way that
electricity
flows through
the circuits of
this device.
Engineers use
their
knowledge of
physics to
construct an

iPhone with features that consumers will enjoy. One specific feature of an iPhone is the GPS function. GPS uses physics equations to determine the driving time between two locations on a map. (credit: @gletham GIS, Social, Mobile Tech Images)

Applications of Physics

You need not be a scientist to use physics. On the contrary, knowledge of physics is useful in everyday situations as well as in nonscientific professions. It can help you understand how microwave ovens work, why metals should not be put into them, and why they might affect pacemakers. (See [\[link\]](#) and [\[link\]](#).) Physics allows you to understand the hazards of radiation and rationally evaluate these hazards more easily. Physics also explains the reason why a black car radiator helps remove heat in a car engine, and it explains why a white roof helps keep the inside of a house cool. Similarly, the operation of a car's ignition system as well as the transmission of electrical signals through our body's nervous system are

much easier to understand when you think about them in terms of basic physics.

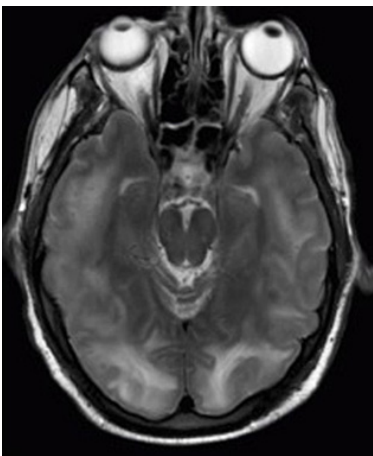
Physics is the foundation of many important disciplines and contributes directly to others. Chemistry, for example—since it deals with the interactions of atoms and molecules—is rooted in atomic and molecular physics. Most branches of engineering are applied physics. In architecture, physics is at the heart of structural stability, and is involved in the acoustics, heating, lighting, and cooling of buildings. Parts of geology rely heavily on physics, such as radioactive dating of rocks, earthquake analysis, and heat transfer in the Earth. Some disciplines, such as biophysics and geophysics, are hybrids of physics and other disciplines.

Physics has many applications in the biological sciences. On the microscopic level, it helps describe the properties of cell walls and cell membranes ([\[link\]](#) and [\[link\]](#)). On the macroscopic level, it can explain the heat, work, and power associated with the human body. Physics is involved in medical diagnostics, such as x-rays, magnetic resonance imaging (MRI), and ultrasonic blood flow measurements. Medical therapy sometimes directly involves physics; for example, cancer radiotherapy uses ionizing radiation. Physics can also explain sensory phenomena, such as how musical instruments make sound, how the eye detects color, and how lasers can transmit information.

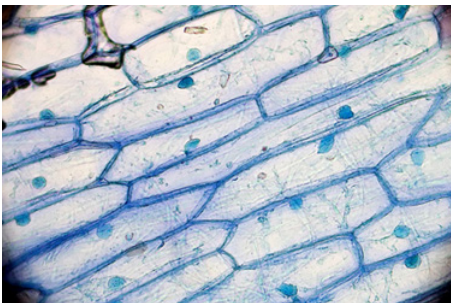
It is not necessary to formally study all applications of physics. What is most useful is knowledge of the basic laws of physics and a skill in the analytical methods for applying them. The study of physics also can improve your problem-solving skills. Furthermore, physics has retained the most basic aspects of science, so it is used by all of the sciences, and the study of physics makes other sciences easier to understand.



The laws of physics help us understand how common appliances work. For example, the laws of physics can help explain how microwave ovens heat up food, and they also help us understand why it is dangerous to place metal objects in a microwave oven. (credit: MoneyBlogNewz)

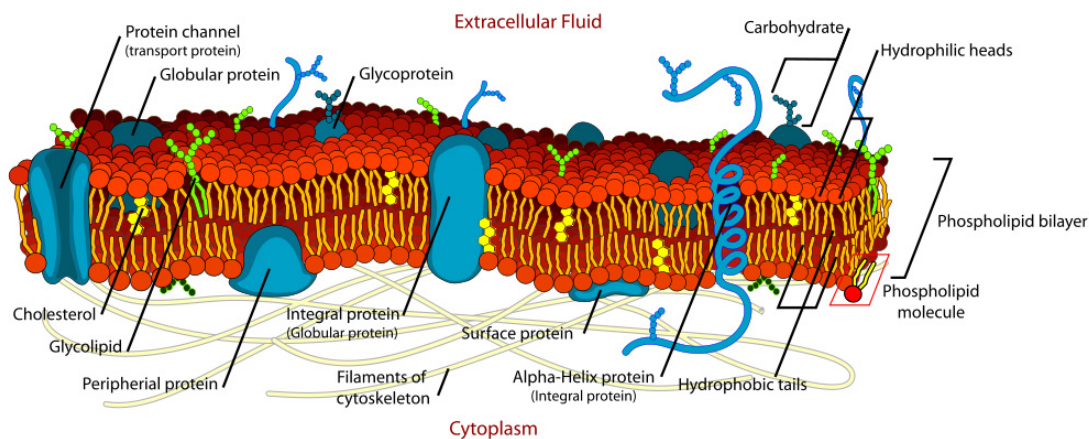


These two applications of physics have more in common than meets the eye. Microwave ovens use electromagnetic waves to heat food. Magnetic resonance imaging (MRI) also uses electromagnetic waves to yield an image of the brain, from which the exact location of tumors can be determined.
(credit: Rashmi Chawla, Daniel Smith, and Paul E. Marik)



Physics, chemistry,

and biology help describe the properties of cell walls in plant cells, such as the onion cells seen here. (credit: Umberto Salvagnin)

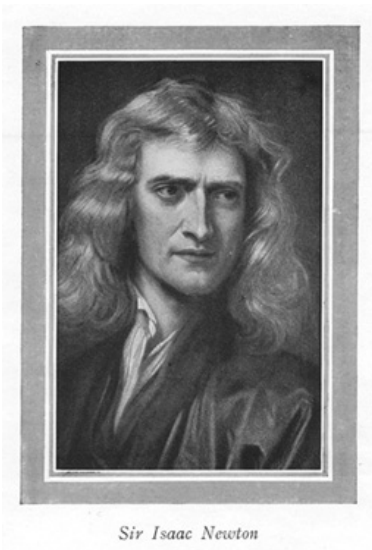


An artist's rendition of the structure of a cell membrane. Membranes form the boundaries of animal cells and are complex in structure and function. Many of the most fundamental properties of life, such as the firing of nerve cells, are related to membranes. The disciplines of biology, chemistry, and physics all help us understand the membranes of animal cells. (credit: Mariana Ruiz)

Models, Theories, and Laws; The Role of Experimentation

The laws of nature are concise descriptions of the universe around us; they are human statements of the underlying laws or rules that all natural processes follow. Such laws are intrinsic to the universe; humans did not

create them and so cannot change them. We can only discover and understand them. Their discovery is a very human endeavor, with all the elements of mystery, imagination, struggle, triumph, and disappointment inherent in any creative effort. (See [\[link\]](#) and [\[link\]](#).) The cornerstone of discovering natural laws is observation; science must describe the universe as it is, not as we may imagine it to be.



Isaac Newton
(1642–1727) was very reluctant to publish his revolutionary work and had to be convinced to do so. In his later years, he stepped down from his academic post and became exchequer of the Royal Mint. He took this post

seriously,
inventing reeding
(or creating
ridges) on the
edge of coins to
prevent
unscrupulous
people from
trimming the
silver off of them
before using them
as currency.
(credit: Arthur
Shuster and
Arthur E. Shipley:
*Britain's Heritage
of Science*.
London, 1917.)



Marie Curie
(1867–1934)
sacrificed

monetary assets
to help finance
her early
research and
damaged her
physical well-
being with
radiation
exposure. She is
the only person
to win Nobel
prizes in both
physics and
chemistry. One
of her daughters
also won a
Nobel Prize.
(credit:
Wikimedia
Commons)

We all are curious to some extent. We look around, make generalizations, and try to understand what we see—for example, we look up and wonder whether one type of cloud signals an oncoming storm. As we become serious about exploring nature, we become more organized and formal in collecting and analyzing data. We attempt greater precision, perform controlled experiments (if we can), and write down ideas about how the data may be organized and unified. We then formulate models, theories, and laws based on the data we have collected and analyzed to generalize and communicate the results of these experiments.

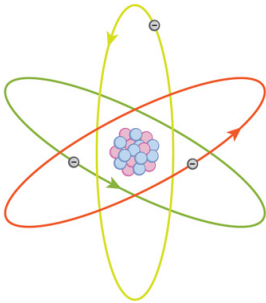
A **model** is a representation of something that is often too difficult (or impossible) to display directly. While a model is justified with experimental proof, it is only accurate under limited situations. An example is the planetary model of the atom in which electrons are pictured as orbiting the

nucleus, analogous to the way planets orbit the Sun. (See [\[link\]](#).) We cannot observe electron orbits directly, but the mental image helps explain the observations we can make, such as the emission of light from hot gases (atomic spectra). Physicists use models for a variety of purposes. For example, models can help physicists analyze a scenario and perform a calculation, or they can be used to represent a situation in the form of a computer simulation. A **theory** is an explanation for patterns in nature that is supported by scientific evidence and verified multiple times by various groups of researchers. Some theories include models to help visualize phenomena, whereas others do not. Newton's theory of gravity, for example, does not require a model or mental image, because we can observe the objects directly with our own senses. The kinetic theory of gases, on the other hand, is a model in which a gas is viewed as being composed of atoms and molecules. Atoms and molecules are too small to be observed directly with our senses—thus, we picture them mentally to understand what our instruments tell us about the behavior of gases.

A **law** uses concise language to describe a generalized pattern in nature that is supported by scientific evidence and repeated experiments. Often, a law can be expressed in the form of a single mathematical equation. Laws and theories are similar in that they are both scientific statements that result from a tested hypothesis and are supported by scientific evidence. However, the designation *law* is reserved for a concise and very general statement that describes phenomena in nature, such as the law that energy is conserved during any process, or Newton's second law of motion, which relates force, mass, and acceleration by the simple equation $\mathbf{F} = m\mathbf{a}$. A theory, in contrast, is a less concise statement of observed phenomena. For example, the Theory of Evolution and the Theory of Relativity cannot be expressed concisely enough to be considered a law. The biggest difference between a law and a theory is that a theory is much more complex and dynamic. A law describes a single action, whereas a theory explains an entire group of related phenomena. And, whereas a law is a postulate that forms the foundation of the scientific method, a theory is the end result of that process.

Less broadly applicable statements are usually called principles (such as Pascal's principle, which is applicable only in fluids), but the distinction

between laws and principles often is not carefully made.



What is a
model?

This
planetary
model of
the atom
shows
electrons
orbiting the
nucleus. It
is a
drawing
that we use
to form a
mental
image of
the atom
that we
cannot see
directly
with our
eyes
because it
is too
small.

Note:**Models, Theories, and Laws**

Models, theories, and laws are used to help scientists analyze the data they have already collected. However, often after a model, theory, or law has been developed, it points scientists toward new discoveries they would not otherwise have made.

The models, theories, and laws we devise sometimes *imply the existence of objects or phenomena as yet unobserved*. These predictions are remarkable triumphs and tributes to the power of science. It is the underlying order in the universe that enables scientists to make such spectacular predictions. However, if *experiment* does not verify our predictions, then the theory or law is wrong, no matter how elegant or convenient it is. Laws can never be known with absolute certainty because it is impossible to perform every imaginable experiment in order to confirm a law in every possible scenario. Physicists operate under the assumption that all scientific laws and theories are valid until a counterexample is observed. If a good-quality, verifiable experiment contradicts a well-established law, then the law must be modified or overthrown completely.

The study of science in general and physics in particular is an adventure much like the exploration of uncharted ocean. Discoveries are made; models, theories, and laws are formulated; and the beauty of the physical universe is made more sublime for the insights gained.

Note:**The Scientific Method**

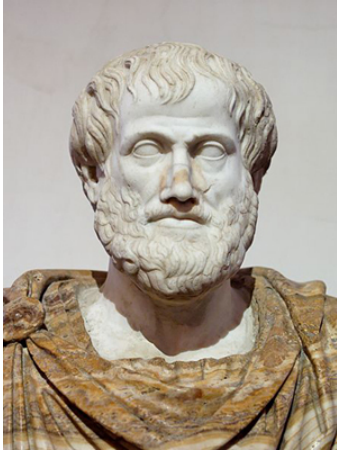
As scientists inquire and gather information about the world, they follow a process called the **scientific method**. This process typically begins with an observation and question that the scientist will research. Next, the scientist

typically performs some research about the topic and then devises a hypothesis. Then, the scientist will test the hypothesis by performing an experiment. Finally, the scientist analyzes the results of the experiment and draws a conclusion. Note that the scientific method can be applied to many situations that are not limited to science, and this method can be modified to suit the situation.

Consider an example. Let us say that you try to turn on your car, but it will not start. You undoubtedly wonder: Why will the car not start? You can follow a scientific method to answer this question. First off, you may perform some research to determine a variety of reasons why the car will not start. Next, you will state a hypothesis. For example, you may believe that the car is not starting because it has no engine oil. To test this, you open the hood of the car and examine the oil level. You observe that the oil is at an acceptable level, and you thus conclude that the oil level is not contributing to your car issue. To troubleshoot the issue further, you may devise a new hypothesis to test and then repeat the process again.

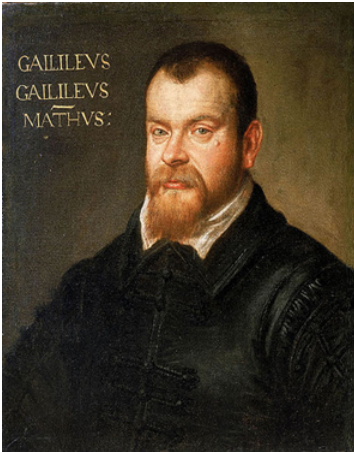
The Evolution of Natural Philosophy into Modern Physics

Physics was not always a separate and distinct discipline. It remains connected to other sciences to this day. The word *physics* comes from Greek, meaning nature. The study of nature came to be called “natural philosophy.” From ancient times through the Renaissance, natural philosophy encompassed many fields, including astronomy, biology, chemistry, physics, mathematics, and medicine. Over the last few centuries, the growth of knowledge has resulted in ever-increasing specialization and branching of natural philosophy into separate fields, with physics retaining the most basic facets. (See [\[link\]](#), [\[link\]](#), and [\[link\]](#).) Physics as it developed from the Renaissance to the end of the 19th century is called **classical physics**. It was transformed into modern physics by revolutionary discoveries made starting at the beginning of the 20th century.



Over the centuries, natural philosophy has evolved into more specialized disciplines, as illustrated by the contributions of some of the greatest minds in history. The Greek philosopher **Aristotle** (384–322 B.C.) wrote on a broad range of topics including physics, animals, the soul, politics, and poetry.
(credit: Jastrow

(2006)/Ludovisi
Collection)



Galileo Galilei
(1564–1642) laid
the foundation of
modern
experimentation
and made
contributions in
mathematics,
physics, and
astronomy.
(credit:
Domenico
Tintoretto)



Niels Bohr
(1885–1962)
made
fundamental
contributions to
the development
of quantum
mechanics, one
part of modern
physics. (credit:
United States
Library of
Congress Prints
and Photographs
Division)

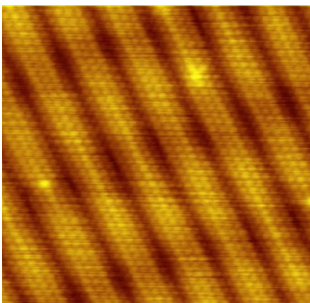
Classical physics is not an exact description of the universe, but it is an excellent approximation under the following conditions: Matter must be moving at speeds less than about 1% of the speed of light, the objects dealt with must be large enough to be seen with a microscope, and only weak gravitational fields, such as the field generated by the Earth, can be involved. Because humans live under such circumstances, classical physics seems intuitively reasonable, while many aspects of modern physics seem bizarre. This is why models are so useful in modern physics—they let us

conceptualize phenomena we do not ordinarily experience. We can relate to models in human terms and visualize what happens when objects move at high speeds or imagine what objects too small to observe with our senses might be like. For example, we can understand an atom's properties because we can picture it in our minds, although we have never seen an atom with our eyes. New tools, of course, allow us to better picture phenomena we cannot see. In fact, new instrumentation has allowed us in recent years to actually “picture” the atom.

Note:

Limits on the Laws of Classical Physics

For the laws of classical physics to apply, the following criteria must be met: Matter must be moving at speeds less than about 1% of the speed of light, the objects dealt with must be large enough to be seen with a microscope, and only weak gravitational fields (such as the field generated by the Earth) can be involved.



Using a
scanning
tunneling
microscope
(STM),
scientists can
see the
individual
atoms that

compose this
sheet of gold.
(credit:
Erwinrossen)

Some of the most spectacular advances in science have been made in modern physics. Many of the laws of classical physics have been modified or rejected, and revolutionary changes in technology, society, and our view of the universe have resulted. Like science fiction, modern physics is filled with fascinating objects beyond our normal experiences, but it has the advantage over science fiction of being very real. Why, then, is the majority of this text devoted to topics of classical physics? There are two main reasons: Classical physics gives an extremely accurate description of the universe under a wide range of everyday circumstances, and knowledge of classical physics is necessary to understand modern physics.

Modern physics itself consists of the two revolutionary theories, relativity and quantum mechanics. These theories deal with the very fast and the very small, respectively. **Relativity** must be used whenever an object is traveling at greater than about 1% of the speed of light or experiences a strong gravitational field such as that near the Sun. **Quantum mechanics** must be used for objects smaller than can be seen with a microscope. The combination of these two theories is *relativistic quantum mechanics*, and it describes the behavior of small objects traveling at high speeds or experiencing a strong gravitational field. Relativistic quantum mechanics is the best universally applicable theory we have. Because of its mathematical complexity, it is used only when necessary, and the other theories are used whenever they will produce sufficiently accurate results. We will find, however, that we can do a great deal of modern physics with the algebra and trigonometry used in this text.

Exercise:

Check Your Understanding

Problem:

A friend tells you he has learned about a new law of nature. What can you know about the information even before your friend describes the law? How would the information be different if your friend told you he had learned about a scientific theory rather than a law?

Solution:

Without knowing the details of the law, you can still infer that the information your friend has learned conforms to the requirements of all laws of nature: it will be a concise description of the universe around us; a statement of the underlying rules that all natural processes follow. If the information had been a theory, you would be able to infer that the information will be a large-scale, broadly applicable generalization.

Note:

PhET Explorations: Equation Grapher

Learn about graphing polynomials. The shape of the curve changes as the constants are adjusted. View the curves for the individual terms (e.g. $y = bx$) to see how they add to generate the polynomial curve.

https://phet.colorado.edu/sims/equation-grapher/equation-grapher_en.html

Summary

- Science seeks to discover and describe the underlying order and simplicity in nature.
- Physics is the most basic of the sciences, concerning itself with energy, matter, space and time, and their interactions.
- Scientific laws and theories express the general truths of nature and the body of knowledge they encompass. These laws of nature are rules that all natural processes appear to follow.

Conceptual Questions

Exercise:

Problem:

Models are particularly useful in relativity and quantum mechanics, where conditions are outside those normally encountered by humans. What is a model?

Exercise:

Problem: How does a model differ from a theory?

Exercise:

Problem:

If two different theories describe experimental observations equally well, can one be said to be more valid than the other (assuming both use accepted rules of logic)?

Exercise:

Problem: What determines the validity of a theory?

Exercise:

Problem:

Certain criteria must be satisfied if a measurement or observation is to be believed. Will the criteria necessarily be as strict for an expected result as for an unexpected result?

Exercise:

Problem:

Can the validity of a model be limited, or must it be universally valid? How does this compare to the required validity of a theory or a law?

Exercise:

Problem:

Classical physics is a good approximation to modern physics under certain circumstances. What are they?

Exercise:

Problem: When is it *necessary* to use relativistic quantum mechanics?

Exercise:**Problem:**

Can classical physics be used to accurately describe a satellite moving at a speed of 7500 m/s? Explain why or why not.

Glossary

classical physics

physics that was developed from the Renaissance to the end of the 19th century

physics

the science concerned with describing the interactions of energy, matter, space, and time; it is especially interested in what fundamental mechanisms underlie every phenomenon

model

representation of something that is often too difficult (or impossible) to display directly

theory

an explanation for patterns in nature that is supported by scientific evidence and verified multiple times by various groups of researchers

law

a description, using concise language or a mathematical formula, a generalized pattern in nature that is supported by scientific evidence

and repeated experiments

scientific method

a method that typically begins with an observation and question that the scientist will research; next, the scientist typically performs some research about the topic and then devises a hypothesis; then, the scientist will test the hypothesis by performing an experiment; finally, the scientist analyzes the results of the experiment and draws a conclusion

modern physics

the study of relativity, quantum mechanics, or both

relativity

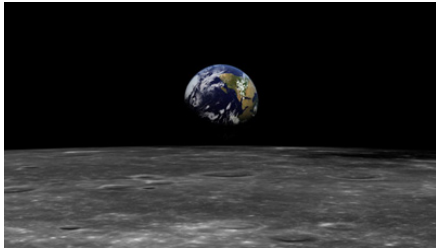
the study of objects moving at speeds greater than about 1% of the speed of light, or of objects being affected by a strong gravitational field

quantum mechanics

the study of objects smaller than can be seen with a microscope

Physical Quantities and Units

- Perform unit conversions both in the SI and English units.
- Explain the most common prefixes in the SI units and be able to write them in scientific notation.

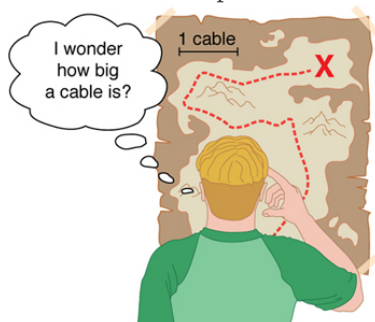


The distance from Earth to the Moon may seem immense, but it is just a tiny fraction of the distances from Earth to other celestial bodies. (credit: NASA)

The range of objects and phenomena studied in physics is immense. From the incredibly short lifetime of a nucleus to the age of the Earth, from the tiny sizes of sub-nuclear particles to the vast distance to the edges of the known universe, from the force exerted by a jumping flea to the force between Earth and the Sun, there are enough factors of 10 to challenge the imagination of even the most experienced scientist. Giving numerical values for physical quantities and equations for physical principles allows us to understand nature much more deeply than does qualitative description alone. To comprehend these vast ranges, we must also have accepted units in which to express them. And we shall find that (even in the potentially mundane discussion of meters, kilograms, and seconds) a profound simplicity of nature appears—all physical quantities can be expressed as combinations of only four fundamental physical quantities: length, mass, time, and electric current.

We define a **physical quantity** either by *specifying how it is measured* or by *stating how it is calculated* from other measurements. For example, we define distance and time by specifying methods for measuring them, whereas we define *average speed* by stating that it is calculated as distance traveled divided by time of travel.

Measurements of physical quantities are expressed in terms of **units**, which are standardized values. For example, the length of a race, which is a physical quantity, can be expressed in units of meters (for sprinters) or kilometers (for distance runners). Without standardized units, it would be extremely difficult for scientists to express and compare measured values in a meaningful way. (See [\[link\]](#).)



Distances given in
unknown units are
maddeningly useless.

There are two major systems of units used in the world: **SI units** (also known as the metric system) and **English units** (also known as the customary or imperial system). **English units** were historically used in nations once ruled by the British Empire and are still widely used in the United States. Virtually every other country in the world now uses SI units as the standard; the metric system is also the standard system agreed upon by scientists and mathematicians. The acronym “SI” is derived from the French *Système International*.

SI Units: Fundamental and Derived Units

[\[link\]](#) gives the fundamental SI units that are used throughout this textbook. This text uses non-SI units in a few applications where they are in very common use, such as the measurement of blood pressure in millimeters of mercury (mm Hg). Whenever non-SI units are discussed, they will be tied to SI units through conversions.

Length	Mass	Time	Electric Current
meter (m)	kilogram (kg)	second (s)	ampere (A)

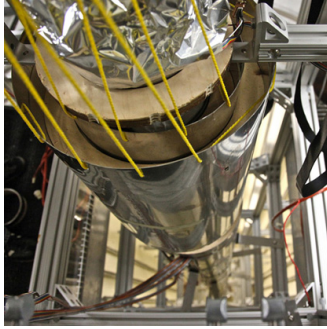
Fundamental SI Units

It is an intriguing fact that some physical quantities are more fundamental than others and that the most fundamental physical quantities can be defined *only* in terms of the procedure used to measure them. The units in which they are measured are thus called **fundamental units**. In this textbook, the fundamental physical quantities are taken to be length, mass, time, and electric current. (Note that electric current will not be introduced until much later in this text.) All other physical quantities, such as force and electric charge, can be expressed as algebraic combinations of length, mass, time, and current (for example, speed is length divided by time); these units are called **derived units**.

Units of Time, Length, and Mass: The Second, Meter, and Kilogram

The Second

The SI unit for time, the **second**(abbreviated s), has a long history. For many years it was defined as 1/86,400 of a mean solar day. More recently, a new standard was adopted to gain greater accuracy and to define the second in terms of a non-varying, or constant, physical phenomenon (because the solar day is getting longer due to very gradual slowing of the Earth’s rotation). Cesium atoms can be made to vibrate in a very steady way, and these vibrations can be readily observed and counted. In 1967 the second was redefined as the time required for 9,192,631,770 of these vibrations. (See [\[link\]](#).) Accuracy in the fundamental units is essential, because all measurements are ultimately expressed in terms of fundamental units and can be no more accurate than are the fundamental units themselves.



An atomic clock such as this one uses the vibrations of cesium atoms to keep time to a precision of better than a microsecond per year. The fundamental unit of time, the second, is based on such clocks. This image is looking down from the top of an atomic fountain nearly 30 feet tall!
(credit: Steve Jurvetson/Flickr)

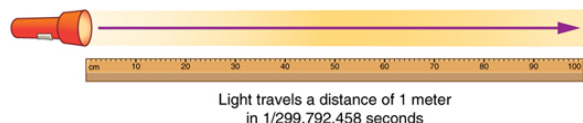
The Meter

The SI unit for length is the **meter** (abbreviated m); its definition has also changed over time to become more accurate and precise. The meter was first defined in 1791 as 1/10,000,000 of the distance from the equator to the North Pole. This measurement was improved in 1889 by redefining the meter to be the distance between two engraved lines on a platinum-iridium bar now kept near Paris. By 1960, it had become possible to define the meter even more accurately in terms of the wavelength of light, so it was again redefined as 1,650,763.73 wavelengths of orange light emitted by krypton atoms. In 1983, the meter was given its present definition (partly for greater accuracy) as the distance light travels in a vacuum in 1/299,792,458 of a second. (See [\[link\]](#).) This change defines the speed of light to be exactly 299,792,458 meters per second. The length of the meter will change if the speed of light is someday measured with greater accuracy.

The Kilogram

The SI unit for mass is the **kilogram** (abbreviated kg); it is defined to be the mass of a platinum-iridium cylinder kept with the old meter standard at the International Bureau of Weights and Measures near Paris. Exact replicas of the standard kilogram are also kept at the United States' National Institute of Standards

and Technology, or NIST, located in Gaithersburg, Maryland outside of Washington D.C., and at other locations around the world. The determination of all other masses can be ultimately traced to a comparison with the standard mass.



The meter is defined to be the distance light travels in $1/299,792,458$ of a second in a vacuum. Distance traveled is speed multiplied by time.

Electric current and its accompanying unit, the ampere, will be introduced in [Introduction to Electric Current, Resistance, and Ohm's Law](#) when electricity and magnetism are covered. The initial modules in this textbook are concerned with mechanics, fluids, heat, and waves. In these subjects all pertinent physical quantities can be expressed in terms of the fundamental units of length, mass, and time.

Metric Prefixes

SI units are part of the **metric system**. The metric system is convenient for scientific and engineering calculations because the units are categorized by factors of 10. [\[link\]](#) gives metric prefixes and symbols used to denote various factors of 10.

Metric systems have the advantage that conversions of units involve only powers of 10. There are 100 centimeters in a meter, 1000 meters in a kilometer, and so on. In nonmetric systems, such as the system of U.S. customary units, the relationships are not as simple—there are 12 inches in a foot, 5280 feet in a mile, and so on. Another advantage of the metric system is that the same unit can be used over extremely large ranges of values simply by using an appropriate metric prefix. For example, distances in meters are suitable in construction, while distances in kilometers are appropriate for air travel, and the tiny measure of nanometers are convenient in optical design. With the metric system there is no need to invent new units for particular applications.

The term **order of magnitude** refers to the scale of a value expressed in the metric system. Each power of 10 in the metric system represents a different order of magnitude. For example, 10^1 , 10^2 , 10^3 , and so forth are all different orders of magnitude. All quantities that can be expressed as a product of a specific power of 10 are said to be of the *same* order of magnitude. For example, the number 800 can be written as 8×10^2 , and the number 450 can be written as 4.5×10^2 . Thus, the numbers 800 and 450 are of the same order of magnitude: 10^2 . Order of magnitude can be thought of as a ballpark estimate for the scale of a value. The diameter of an atom is on the order of 10^{-9} m, while the diameter of the Sun is on the order of 10^9 m.

Note:

The Quest for Microscopic Standards for Basic Units

The fundamental units described in this chapter are those that produce the greatest accuracy and precision in measurement. There is a sense among physicists that, because there is an underlying microscopic substructure to matter, it would be most satisfying to base our standards of measurement on microscopic objects and fundamental physical phenomena such as the speed of light. A microscopic standard has been accomplished for the standard of time, which is based on the oscillations of the cesium atom.

The standard for length was once based on the wavelength of light (a small-scale length) emitted by a certain type of atom, but it has been supplanted by the more precise measurement of the speed of light. If it becomes possible to measure the mass of atoms or a particular arrangement of atoms such as a silicon sphere to greater precision than the kilogram standard, it may become possible to base mass measurements on the small scale. There are also possibilities that electrical phenomena on the small scale may someday allow us to base a unit of charge on the charge of electrons and protons, but at present current and charge are related to large-scale currents and forces between wires.

Prefix	Symbol	Value ^[footnote] See Appendix A for a discussion of powers of 10.	Example (some are approximate)			
exa	E	10^{18}	exameter	Em	10^{18} m	distance light travels in a century
peta	P	10^{15}	petasecond	Ps	10^{15} s	30 million years
tera	T	10^{12}	terawatt	TW	10^{12} W	powerful laser output
giga	G	10^9	gigahertz	GHz	10^9 Hz	a microwave frequency
mega	M	10^6	megacurie	MCi	10^6 Ci	high radioactivity
kilo	k	10^3	kilometer	km	10^3 m	about 6/10 mile
hecto	h	10^2	hectoliter	hL	10^2 L	26 gallons

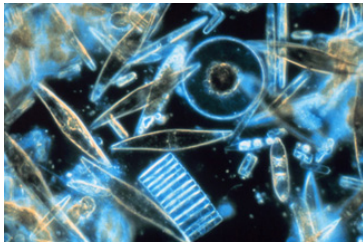
Prefix	Symbol	Value ^{[footnote]} See Appendix A for a discussion of powers of 10.	Example (some are approximate)			
deka	da	10^1	dekagram	dag	10^1 g	teaspoon of butter
—	—	10^0 (=1)				
deci	d	10^{-1}	deciliter	dL	10^{-1} L	less than half a soda
centi	c	10^{-2}	centimeter	cm	10^{-2} m	fingertip thickness
milli	m	10^{-3}	millimeter	mm	10^{-3} m	flea at its shoulders
micro	μ	10^{-6}	micrometer	μm	10^{-6} m	detail in microscope
nano	n	10^{-9}	nanogram	ng	10^{-9} g	small speck of dust
pico	p	10^{-12}	picofarad	pF	10^{-12} F	small capacitor in radio
femto	f	10^{-15}	femtometer	fm	10^{-15} m	size of a proton
atto	a	10^{-18}	attosecond	as	10^{-18} s	time light crosses an atom

Metric Prefixes for Powers of 10 and their Symbols

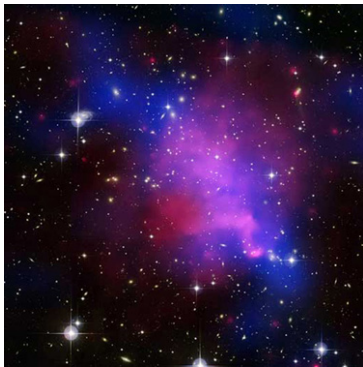
Known Ranges of Length, Mass, and Time

The vastness of the universe and the breadth over which physics applies are illustrated by the wide range of examples of known lengths, masses, and times in [\[link\]](#). Examination of this table will give you some

feeling for the range of possible topics and numerical values. (See [\[link\]](#) and [\[link\]](#).)



Tiny phytoplankton
swims among crystals of
ice in the Antarctic Sea.
They range from a few
micrometers to as much
as 2 millimeters in length.
(credit: Prof. Gordon T.
Taylor, Stony Brook
University; NOAA Corps
Collections)



Galaxies collide 2.4
billion light years away
from Earth. The
tremendous range of
observable phenomena in
nature challenges the
imagination. (credit:
NASA/CXC/UVic./A.
Mahdavi et al.
Optical/lensing:
CFHT/UVic./H. Hoekstra
et al.)

Unit Conversion and Dimensional Analysis

It is often necessary to convert from one type of unit to another. For example, if you are reading a European cookbook, some quantities may be expressed in units of liters and you need to convert them to cups. Or, perhaps you are reading walking directions from one location to another and you are interested in how many miles you will be walking. In this case, you will need to convert units of feet to miles.

Let us consider a simple example of how to convert units. Let us say that we want to convert 80 meters (m) to kilometers (km).

The first thing to do is to list the units that you have and the units that you want to convert to. In this case, we have units in *meters* and we want to convert to *kilometers*.

Next, we need to determine a **conversion factor** relating meters to kilometers. A conversion factor is a ratio expressing how many of one unit are equal to another unit. For example, there are 12 inches in 1 foot, 100 centimeters in 1 meter, 60 seconds in 1 minute, and so on. In this case, we know that there are 1,000 meters in 1 kilometer.

Now we can set up our unit conversion. We will write the units that we have and then multiply them by the conversion factor so that the units cancel out, as shown:

Equation:

80 m × (1 km / 1000 m) = 0.080 km.

Note that the unwanted m unit cancels, leaving only the desired km unit. You can use this method to convert between any types of unit.

Click [\[link\]](#) for a more complete list of conversion factors.

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10 ⁻¹⁸	Present experimental limit to smallest observable detail	10 ⁻³⁰	Mass of an electron (9.11 × 10 ⁻³¹ kg)	10 ⁻²³	Time for light to cross a proton
10 ⁻¹⁵	Diameter of a proton	10 ⁻²⁷	Mass of a hydrogen atom (1.67 × 10 ⁻²⁷ kg)	10 ⁻²²	Mean life of an extremely unstable nucleus

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10^{-14}	Diameter of a uranium nucleus	10^{-15}	Mass of a bacterium	10^{-15}	Time for one oscillation of visible light
10^{-10}	Diameter of a hydrogen atom	10^{-5}	Mass of a mosquito	10^{-13}	Time for one vibration of an atom in a solid
10^{-8}	Thickness of membranes in cells of living organisms	10^{-2}	Mass of a hummingbird	10^{-8}	Time for one oscillation of an FM radio wave
10^{-6}	Wavelength of visible light	1	Mass of a liter of water (about a quart)	10^{-3}	Duration of a nerve impulse
10^{-3}	Size of a grain of sand	10^2	Mass of a person	1	Time for one heartbeat
1	Height of a 4-year-old child	10^3	Mass of a car	10^5	One day (8.64×10^4 s)
10^2	Length of a football field	10^8	Mass of a large ship	10^7	One year (y) (3.16×10^7 s)
10^4	Greatest ocean depth	10^{12}	Mass of a large iceberg	10^9	About half the life expectancy of a human
10^7	Diameter of the Earth	10^{15}	Mass of the nucleus of a comet	10^{11}	Recorded history
10^{11}	Distance from the Earth to the Sun	10^{23}	Mass of the Moon (7.35×10^{22} kg)	10^{17}	Age of the Earth
10^{16}	Distance traveled by light in 1 year (a light year)	10^{25}	Mass of the Earth (5.97×10^{24} kg)	10^{18}	Age of the universe
10^{21}	Diameter of the Milky Way galaxy	10^{30}	Mass of the Sun (1.99×10^{30} kg)		

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10^{22}	Distance from the Earth to the nearest large galaxy (Andromeda)	10^{42}	Mass of the Milky Way galaxy (current upper limit)		
10^{26}	Distance from the Earth to the edges of the known universe	10^{53}	Mass of the known universe (current upper limit)		

Approximate Values of Length, Mass, and Time

Example:

Unit Conversions: A Short Drive Home

Suppose that you drive the 10.0 km from your university to home in 20.0 min. Calculate your average speed (a) in kilometers per hour (km/h) and (b) in meters per second (m/s). (Note: Average speed is distance traveled divided by time of travel.)

Strategy

First we calculate the average speed using the given units. Then we can get the average speed into the desired units by picking the correct conversion factor and multiplying by it. The correct conversion factor is the one that cancels the unwanted unit and leaves the desired unit in its place.

Solution for (a)

(1) Calculate average speed. Average speed is distance traveled divided by time of travel. (Take this definition as a given for now—average speed and other motion concepts will be covered in a later module.) In equation form,

Equation:

$$\text{average speed} = \frac{\text{distance}}{\text{time}}.$$

(2) Substitute the given values for distance and time.

Equation:

$$\text{average speed} = \frac{10.0 \text{ km}}{20.0 \text{ min}} = 0.500 \frac{\text{km}}{\text{min}}.$$

(3) Convert km/min to km/h: multiply by the conversion factor that will cancel minutes and leave hours. That conversion factor is 60 min/hr. Thus,

Equation:

$$\text{average speed} = 0.500 \frac{\text{km}}{\text{min}} \times \frac{60 \text{ min}}{1 \text{ h}} = 30.0 \frac{\text{km}}{\text{h}}.$$

Discussion for (a)

To check your answer, consider the following:

(1) Be sure that you have properly cancelled the units in the unit conversion. If you have written the unit conversion factor upside down, the units will not cancel properly in the equation. If you accidentally get the ratio upside down, then the units will not cancel; rather, they will give you the wrong units as follows:

Equation:

$$\frac{\text{km}}{\text{min}} \times \frac{1 \text{ hr}}{60 \text{ min}} = \frac{1}{60} \frac{\text{km} \cdot \text{hr}}{\text{min}^2},$$

which are obviously not the desired units of km/h.

(2) Check that the units of the final answer are the desired units. The problem asked us to solve for average speed in units of km/h and we have indeed obtained these units.

(3) Check the significant figures. Because each of the values given in the problem has three significant figures, the answer should also have three significant figures. The answer 30.0 km/hr does indeed have three significant figures, so this is appropriate. Note that the significant figures in the conversion factor are not relevant because an hour is *defined* to be 60 minutes, so the precision of the conversion factor is perfect.

(4) Next, check whether the answer is reasonable. Let us consider some information from the problem—if you travel 10 km in a third of an hour (20 min), you would travel three times that far in an hour. The answer does seem reasonable.

Solution for (b)

There are several ways to convert the average speed into meters per second.

(1) Start with the answer to (a) and convert km/h to m/s. Two conversion factors are needed—one to convert hours to seconds, and another to convert kilometers to meters.

(2) Multiplying by these yields

Equation:

$$\text{Average speed} = 30.0 \frac{\text{km}}{\text{h}} \times \frac{1 \text{ h}}{3,600 \text{ s}} \times \frac{1,000 \text{ m}}{1 \text{ km}},$$

Equation:

$$\text{Average speed} = 8.33 \frac{\text{m}}{\text{s}}.$$

Discussion for (b)

If we had started with 0.500 km/min, we would have needed different conversion factors, but the answer would have been the same: 8.33 m/s.

You may have noted that the answers in the worked example just covered were given to three digits.

Why? When do you need to be concerned about the number of digits in something you calculate? Why not write down all the digits your calculator produces? The module [Accuracy, Precision, and Significant Figures](#) will help you answer these questions.

Note:

Nonstandard Units

While there are numerous types of units that we are all familiar with, there are others that are much more obscure. For example, a **firkin** is a unit of volume that was once used to measure beer. One firkin equals about 34 liters. To learn more about nonstandard units, use a dictionary or encyclopedia to research different “weights and measures.” Take note of any unusual units, such as a barleycorn, that are not listed in the text. Think about how the unit is defined and state its relationship to SI units.

Exercise:
Check Your Understanding

Problem:

Some hummingbirds beat their wings more than 50 times per second. A scientist is measuring the time it takes for a hummingbird to beat its wings once. Which fundamental unit should the scientist use to describe the measurement? Which factor of 10 is the scientist likely to use to describe the motion precisely? Identify the metric prefix that corresponds to this factor of 10.

Solution:

The scientist will measure the time between each movement using the fundamental unit of seconds. Because the wings beat so fast, the scientist will probably need to measure in milliseconds, or 10^{-3} seconds. (50 beats per second corresponds to 20 milliseconds per beat.)

Exercise:
Check Your Understanding

Problem:

One cubic centimeter is equal to one milliliter. What does this tell you about the different units in the SI metric system?

Solution:

The fundamental unit of length (meter) is probably used to create the derived unit of volume (liter). The measure of a milliliter is dependent on the measure of a centimeter.

Summary

- Physical quantities are a characteristic or property of an object that can be measured or calculated from other measurements.
- Units are standards for expressing and comparing the measurement of physical quantities. All units can be expressed as combinations of four fundamental units.
- The four fundamental units we will use in this text are the meter (for length), the kilogram (for mass), the second (for time), and the ampere (for electric current). These units are part of the metric system, which uses powers of 10 to relate quantities over the vast ranges encountered in nature.
- The four fundamental units are abbreviated as follows: meter, m; kilogram, kg; second, s; and ampere, A. The metric system also uses a standard set of prefixes to denote each order of magnitude greater than or lesser than the fundamental unit itself.
- Unit conversions involve changing a value expressed in one type of unit to another type of unit. This is done by using conversion factors, which are ratios relating equal quantities of different units.

Conceptual Questions

Exercise:

Problem: Identify some advantages of metric units.

Problems & Exercises

Exercise:**Problem:**

The speed limit on some interstate highways is roughly 100 km/h. (a) What is this in meters per second? (b) How many miles per hour is this?

Solution:

- a. 27.8 m/s
- b. 62.1 mph

Exercise:**Problem:**

A car is traveling at a speed of 33 m/s. (a) What is its speed in kilometers per hour? (b) Is it exceeding the 90 km/h speed limit?

Exercise:**Problem:**

Show that $1.0 \text{ m/s} = 3.6 \text{ km/h}$. Hint: Show the explicit steps involved in converting $1.0 \text{ m/s} = 3.6 \text{ km/h}$.

Solution:

$$\begin{aligned}\frac{1.0 \text{ m}}{\text{s}} &= \frac{1.0 \text{ m}}{\text{s}} \times \frac{3600 \text{ s}}{1 \text{ hr}} \times \frac{1 \text{ km}}{1000 \text{ m}} \\ &= 3.6 \text{ km/h.}\end{aligned}$$

Exercise:**Problem:**

American football is played on a 100-yd-long field, excluding the end zones. How long is the field in meters? (Assume that 1 meter equals 3.281 feet.)

Exercise:**Problem:**

Soccer fields vary in size. A large soccer field is 115 m long and 85 m wide. What are its dimensions in feet and inches? (Assume that 1 meter equals 3.281 feet.)

Solution:

length: 377 ft; 4.53×10^3 in. width: 280 ft; 3.3×10^3 in.

Exercise:**Problem:**

What is the height in meters of a person who is 6 ft 1.0 in. tall? (Assume that 1 meter equals 39.37 in.)

Exercise:

Problem:

Mount Everest, at 29,028 feet, is the tallest mountain on the Earth. What is its height in kilometers? (Assume that 1 kilometer equals 3,281 feet.)

Solution:

8.847 km

Exercise:

Problem: The speed of sound is measured to be 342 m/s on a certain day. What is this in km/h?

Exercise:**Problem:**

Tectonic plates are large segments of the Earth's crust that move slowly. Suppose that one such plate has an average speed of 4.0 cm/year. (a) What distance does it move in 1 s at this speed? (b) What is its speed in kilometers per million years?

Solution:

(a) 1.3×10^{-9} m

(b) 40 km/My

Exercise:**Problem:**

(a) Refer to [\[link\]](#) to determine the average distance between the Earth and the Sun. Then calculate the average speed of the Earth in its orbit in kilometers per second. (b) What is this in meters per second?

Glossary

physical quantity

a characteristic or property of an object that can be measured or calculated from other measurements

units

a standard used for expressing and comparing measurements

SI units

the international system of units that scientists in most countries have agreed to use; includes units such as meters, liters, and grams

English units

system of measurement used in the United States; includes units of measurement such as feet, gallons, and pounds

fundamental units

units that can only be expressed relative to the procedure used to measure them

derived units

units that can be calculated using algebraic combinations of the fundamental units

second

the SI unit for time, abbreviated (s)

meter

the SI unit for length, abbreviated (m)

kilogram

the SI unit for mass, abbreviated (kg)

metric system

a system in which values can be calculated in factors of 10

order of magnitude

refers to the size of a quantity as it relates to a power of 10

conversion factor

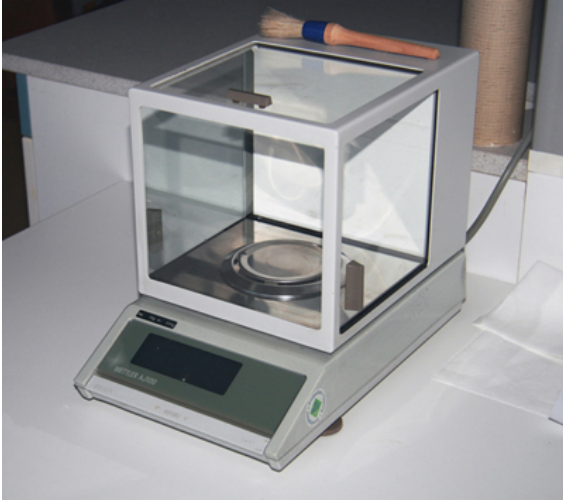
a ratio expressing how many of one unit are equal to another unit

Accuracy, Precision, and Significant Figures

- Determine the appropriate number of significant figures in both addition and subtraction, as well as multiplication and division calculations.
- Calculate the percent uncertainty of a measurement.



A double-pan mechanical balance is used to compare different masses. Usually an object with unknown mass is placed in one pan and objects of known mass are placed in the other pan. When the bar that connects the two pans is horizontal, then the masses in both pans are equal. The “known masses” are typically metal cylinders of standard mass such as 1 gram, 10 grams, and 100 grams.
(credit: Serge Melki)



Many mechanical balances, such as double-pan balances, have been replaced by digital scales, which can typically measure the mass of an object more precisely. Whereas a mechanical balance may only read the mass of an object to the nearest tenth of a gram, many digital scales can measure the mass of an object up to the nearest thousandth of a gram. (credit: Karel Jakubec)

Accuracy and Precision of a Measurement

Science is based on observation and experiment—that is, on measurements. **Accuracy** is how close a measurement is to the correct value for that measurement. For example, let us say that you are measuring the length of standard computer paper. The packaging in which you purchased the paper states that it is 11.0 inches long. You measure the length of the paper three times and obtain the following measurements: 11.1 in., 11.2 in., and 10.9 in.

These measurements are quite accurate because they are very close to the correct value of 11.0 inches. In contrast, if you had obtained a measurement of 12 inches, your measurement would not be very accurate.

The **precision** of a measurement system refers to how close the agreement is between repeated measurements (which are repeated under the same conditions). Consider the example of the paper measurements. The precision of the measurements refers to the spread of the measured values. One way to analyze the precision of the measurements would be to determine the range, or difference, between the lowest and the highest measured values. In that case, the lowest value was 10.9 in. and the highest value was 11.2 in. Thus, the measured values deviated from each other by at most 0.3 in. These measurements were relatively precise because they did not vary too much in value. However, if the measured values had been 10.9, 11.1, and 11.9, then the measurements would not be very precise because there would be significant variation from one measurement to another.

The measurements in the paper example are both accurate and precise, but in some cases, measurements are accurate but not precise, or they are precise but not accurate. Let us consider an example of a GPS system that is attempting to locate the position of a restaurant in a city. Think of the restaurant location as existing at the center of a bull's-eye target, and think of each GPS attempt to locate the restaurant as a black dot. In [\[link\]](#), you can see that the GPS measurements are spread out far apart from each other, but they are all relatively close to the actual location of the restaurant at the center of the target. This indicates a low precision, high accuracy measuring system. However, in [\[link\]](#), the GPS measurements are concentrated quite closely to one another, but they are far away from the target location. This indicates a high precision, low accuracy measuring system.



A GPS system attempts to locate a restaurant at the center of the bull's-eye. The black dots represent each attempt to pinpoint the location of the restaurant. The dots are spread out quite far apart from one another, indicating low precision, but they are each rather close to the actual location of the restaurant, indicating high accuracy.
(credit: Dark Evil)



In this figure,
the dots are
concentrated
rather closely to
one another,
indicating high
precision, but
they are rather
far away from
the actual
location of the
restaurant,
indicating low
accuracy.
(credit: Dark
Evil)

Accuracy, Precision, and Uncertainty

The degree of accuracy and precision of a measuring system are related to the **uncertainty** in the measurements. Uncertainty is a quantitative measure of how much your measured values deviate from a standard or expected value. If your measurements are not very accurate or precise, then the

uncertainty of your values will be very high. In more general terms, uncertainty can be thought of as a disclaimer for your measured values. For example, if someone asked you to provide the mileage on your car, you might say that it is 45,000 miles, plus or minus 500 miles. The plus or minus amount is the uncertainty in your value. That is, you are indicating that the actual mileage of your car might be as low as 44,500 miles or as high as 45,500 miles, or anywhere in between. All measurements contain some amount of uncertainty. In our example of measuring the length of the paper, we might say that the length of the paper is 11 in., plus or minus 0.2 in. The uncertainty in a measurement, A , is often denoted as δA (“delta A ”), so the measurement result would be recorded as $A \pm \delta A$. In our paper example, the length of the paper could be expressed as $11 \text{ in.} \pm 0.2$.

The factors contributing to uncertainty in a measurement include:

1. Limitations of the measuring device,
2. The skill of the person making the measurement,
3. Irregularities in the object being measured,
4. Any other factors that affect the outcome (highly dependent on the situation).

In our example, such factors contributing to the uncertainty could be the following: the smallest division on the ruler is 0.1 in., the person using the ruler has bad eyesight, or one side of the paper is slightly longer than the other. At any rate, the uncertainty in a measurement must be based on a careful consideration of all the factors that might contribute and their possible effects.

Note:

Making Connections: Real-World Connections – Fevers or Chills?

Uncertainty is a critical piece of information, both in physics and in many other real-world applications. Imagine you are caring for a sick child. You suspect the child has a fever, so you check his or her temperature with a thermometer. What if the uncertainty of the thermometer were 3.0°C ? If the child’s temperature reading was 37.0°C (which is normal body temperature), the “true” temperature could be anywhere from a

hypothermic 34.0°C to a dangerously high 40.0°C. A thermometer with an uncertainty of 3.0°C would be useless.

Percent Uncertainty

One method of expressing uncertainty is as a percent of the measured value. If a measurement A is expressed with uncertainty, δA , the **percent uncertainty** (%unc) is defined to be

Equation:

$$\% \text{ unc} = \frac{\delta A}{A} \times 100\%.$$

Example:

Calculating Percent Uncertainty: A Bag of Apples

A grocery store sells 5-lb bags of apples. You purchase four bags over the course of a month and weigh the apples each time. You obtain the following measurements:

Week 1 weight: 4.8 lb

Week 2 weight: 5.3 lb

Week 3 weight: 4.9 lb

Week 4 weight: 5.4 lb

You determine that the weight of the 5-lb bag has an uncertainty of ± 0.4 lb. What is the percent uncertainty of the bag's weight?

Strategy

First, observe that the expected value of the bag's weight, A , is 5 lb. The uncertainty in this value, δA , is 0.4 lb. We can use the following equation to determine the percent uncertainty of the weight:

Equation:

$$\% \text{ unc} = \frac{\delta A}{A} \times 100\%.$$

Solution

Plug the known values into the equation:

Equation:

$$\% \text{ unc} = \frac{0.4 \text{ lb}}{5 \text{ lb}} \times 100\% = 8\%.$$

Discussion

We can conclude that the weight of the apple bag is $5 \text{ lb} \pm 8\%$. Consider how this percent uncertainty would change if the bag of apples were half as heavy, but the uncertainty in the weight remained the same. Hint for future calculations: when calculating percent uncertainty, always remember that you must multiply the fraction by 100%. If you do not do this, you will have a decimal quantity, not a percent value.

Uncertainties in Calculations

There is an uncertainty in anything calculated from measured quantities. For example, the area of a floor calculated from measurements of its length and width has an uncertainty because the length and width have uncertainties. How big is the uncertainty in something you calculate by multiplication or division? If the measurements going into the calculation have small uncertainties (a few percent or less), then the **method of adding percents** can be used for multiplication or division. This method says that *the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation*. For example, if a floor has a length of 4.00 m and a width of 3.00 m, with uncertainties of 2% and 1%, respectively, then the area of the floor is 12.0 m^2 and has an uncertainty of 3%. (Expressed as an area this is 0.36 m^2 , which we round to 0.4 m^2 since the area of the floor is given to a tenth of a square meter.)

Exercise:

Check Your Understanding

Problem:

A high school track coach has just purchased a new stopwatch. The stopwatch manual states that the stopwatch has an uncertainty of ± 0.05 s. Runners on the track coach's team regularly clock 100-m sprints of 11.49 s to 15.01 s. At the school's last track meet, the first-place sprinter came in at 12.04 s and the second-place sprinter came in at 12.07 s. Will the coach's new stopwatch be helpful in timing the sprint team? Why or why not?

Solution:

No, the uncertainty in the stopwatch is too great to effectively differentiate between the sprint times.

Precision of Measuring Tools and Significant Figures

An important factor in the accuracy and precision of measurements involves the precision of the measuring tool. In general, a precise measuring tool is one that can measure values in very small increments. For example, a standard ruler can measure length to the nearest millimeter, while a caliper can measure length to the nearest 0.01 millimeter. The caliper is a more precise measuring tool because it can measure extremely small differences in length. The more precise the measuring tool, the more precise and accurate the measurements can be.

When we express measured values, we can only list as many digits as we initially measured with our measuring tool. For example, if you use a standard ruler to measure the length of a stick, you may measure it to be 36.7 cm. You could not express this value as 36.71 cm because your measuring tool was not precise enough to measure a hundredth of a centimeter. It should be noted that the last digit in a measured value has been estimated in some way by the person performing the measurement. For example, the person measuring the length of a stick with a ruler notices that the stick length seems to be somewhere in between 36.6 cm and 36.7 cm, and he or she must estimate the value of the last digit. Using the

method of **significant figures**, the rule is that *the last digit written down in a measurement is the first digit with some uncertainty*. In order to determine the number of significant digits in a value, start with the first measured value at the left and count the number of digits through the last digit written on the right. For example, the measured value 36.7 cm has three digits, or significant figures. Significant figures indicate the precision of a measuring tool that was used to measure a value.

Zeros

Special consideration is given to zeros when counting significant figures. The zeros in 0.053 are not significant, because they are only placekeepers that locate the decimal point. There are two significant figures in 0.053. The zeros in 10.053 are not placekeepers but are significant—this number has five significant figures. The zeros in 1300 may or may not be significant depending on the style of writing numbers. They could mean the number is known to the last digit, or they could be placekeepers. So 1300 could have two, three, or four significant figures. (To avoid this ambiguity, write 1300 in scientific notation.) *Zeros are significant except when they serve only as placekeepers.*

Exercise:

Check Your Understanding

Problem:

Determine the number of significant figures in the following measurements:

- a. 0.0009
- b. 15,450.0
- c. 6×10^3
- d. 87.990
- e. 30.42

Solution:

- (a) 1; the zeros in this number are placekeepers that indicate the decimal point
- (b) 6; here, the zeros indicate that a measurement was made to the 0.1 decimal point, so the zeros are significant
- (c) 1; the value 10^3 signifies the decimal place, not the number of measured values
- (d) 5; the final zero indicates that a measurement was made to the 0.001 decimal point, so it is significant
- (e) 4; any zeros located in between significant figures in a number are also significant

Significant Figures in Calculations

When combining measurements with different degrees of accuracy and precision, *the number of significant digits in the final answer can be no greater than the number of significant digits in the least precise measured value*. There are two different rules, one for multiplication and division and the other for addition and subtraction, as discussed below.

1. For multiplication and division: *The result should have the same number of significant figures as the quantity having the least significant figures entering into the calculation.* For example, the area of a circle can be calculated from its radius using $A = \pi r^2$. Let us see how many significant figures the area has if the radius has only two—say, $r = 1.2$ m. Then,

Equation:

$$A = \pi r^2 = (3.1415927...) \times (1.2 \text{ m})^2 = 4.5238934 \text{ m}^2$$

is what you would get using a calculator that has an eight-digit output. But because the radius has only two significant figures, it limits the calculated

quantity to two significant figures or

Equation:

$$A=4.5 \text{ m}^2,$$

even though π is good to at least eight digits.

2. For addition and subtraction: *The answer can contain no more decimal places than the least precise measurement.* Suppose that you buy 7.56-kg of potatoes in a grocery store as measured with a scale with precision 0.01 kg. Then you drop off 6.052-kg of potatoes at your laboratory as measured by a scale with precision 0.001 kg. Finally, you go home and add 13.7 kg of potatoes as measured by a bathroom scale with precision 0.1 kg. How many kilograms of potatoes do you now have, and how many significant figures are appropriate in the answer? The mass is found by simple addition and subtraction:

Equation:

$$\begin{array}{r} 7.56 \text{ kg} \\ - 6.052 \text{ kg} \\ \hline +13.7 \text{ kg} \\ 15.208 \text{ kg} \end{array} = 15.2 \text{ kg}.$$

Next, we identify the least precise measurement: 13.7 kg. This measurement is expressed to the 0.1 decimal place, so our final answer must also be expressed to the 0.1 decimal place. Thus, the answer is rounded to the tenths place, giving us 15.2 kg.

Significant Figures in this Text

In this text, most numbers are assumed to have three significant figures. Furthermore, consistent numbers of significant figures are used in all worked examples. You will note that an answer given to three digits is based on input good to at least three digits, for example. If the input has fewer significant figures, the answer will also have fewer significant

figures. Care is also taken that the number of significant figures is reasonable for the situation posed. In some topics, particularly in optics, more accurate numbers are needed and more than three significant figures will be used. Finally, if a number is *exact*, such as the two in the formula for the circumference of a circle, $c = 2\pi r$, it does not affect the number of significant figures in a calculation.

Exercise:

Check Your Understanding

Problem:

Perform the following calculations and express your answer using the correct number of significant digits.

- (a) A woman has two bags weighing 13.5 pounds and one bag with a weight of 10.2 pounds. What is the total weight of the bags?
- (b) The force F on an object is equal to its mass m multiplied by its acceleration a . If a wagon with mass 55 kg accelerates at a rate of 0.0255 m/s^2 , what is the force on the wagon? (The unit of force is called the newton, and it is expressed with the symbol N.)

Solution:

- (a) 37.2 pounds; Because the number of bags is an exact value, it is not considered in the significant figures.
- (b) 1.4 N; Because the value 55 kg has only two significant figures, the final value must also contain two significant figures.

Note:

PhET Explorations: Estimation

Explore size estimation in one, two, and three dimensions! Multiple levels of difficulty allow for progressive skill improvement.

https://phet.colorado.edu/sims/estimation/estimation_en.html

Summary

- Accuracy of a measured value refers to how close a measurement is to the correct value. The uncertainty in a measurement is an estimate of the amount by which the measurement result may differ from this value.
- Precision of measured values refers to how close the agreement is between repeated measurements.
- The precision of a *measuring tool* is related to the size of its measurement increments. The smaller the measurement increment, the more precise the tool.
- Significant figures express the precision of a measuring tool.
- When multiplying or dividing measured values, the final answer can contain only as many significant figures as the least precise value.
- When adding or subtracting measured values, the final answer cannot contain more decimal places than the least precise value.

Conceptual Questions

Exercise:

Problem:

What is the relationship between the accuracy and uncertainty of a measurement?

Exercise:

Problem:

Prescriptions for vision correction are given in units called *diopters* (D). Determine the meaning of that unit. Obtain information (perhaps by calling an optometrist or performing an internet search) on the minimum uncertainty with which corrections in diopters are determined and the accuracy with which corrective lenses can be produced. Discuss the sources of uncertainties in both the prescription and accuracy in the manufacture of lenses.

Problems & Exercises

Express your answers to problems in this section to the correct number of significant figures and proper units.

Exercise:

Problem:

Suppose that your bathroom scale reads your mass as 65 kg with a 3% uncertainty. What is the uncertainty in your mass (in kilograms)?

Solution:

2 kg

Exercise:

Problem:

A good-quality measuring tape can be off by 0.50 cm over a distance of 20 m. What is its percent uncertainty?

Exercise:

Problem:

(a) A car speedometer has a 5.0% uncertainty. What is the range of possible speeds when it reads 90 km/h? (b) Convert this range to miles per hour. (1 km = 0.6214 mi)

Solution:

a. 85.5 to 94.5 km/h

b. 53.1 to 58.7 mi/h

Exercise:

Problem:

An infant's pulse rate is measured to be 130 ± 5 beats/min. What is the percent uncertainty in this measurement?

Exercise:**Problem:**

(a) Suppose that a person has an average heart rate of 72.0 beats/min. How many beats does he or she have in 2.0 y? (b) In 2.00 y? (c) In 2.000 y?

Solution:

(a) 7.6×10^7 beats

(b) 7.57×10^7 beats

(c) 7.57×10^7 beats

Exercise:**Problem:**

A can contains 375 mL of soda. How much is left after 308 mL is removed?

Exercise:**Problem:**

State how many significant figures are proper in the results of the following calculations: (a) $(106.7)(98.2)/(46.210)(1.01)$ (b) $(18.7)^2$ (c) $(1.60 \times 10^{-19})(3712)$.

Solution:

a. 3

b. 3

c. 3

Exercise:

Problem:

(a) How many significant figures are in the numbers 99 and 100? (b) If the uncertainty in each number is 1, what is the percent uncertainty in each? (c) Which is a more meaningful way to express the accuracy of these two numbers, significant figures or percent uncertainties?

Exercise:**Problem:**

(a) If your speedometer has an uncertainty of 2.0 km/h at a speed of 90 km/h, what is the percent uncertainty? (b) If it has the same percent uncertainty when it reads 60 km/h, what is the range of speeds you could be going?

Solution:

a) 2.2%

(b) 59 to 61 km/h

Exercise:**Problem:**

(a) A person's blood pressure is measured to be 120 ± 2 mm Hg. What is its percent uncertainty? (b) Assuming the same percent uncertainty, what is the uncertainty in a blood pressure measurement of 80 mm Hg?

Exercise:**Problem:**

A person measures his or her heart rate by counting the number of beats in 30 s. If 40 ± 1 beats are counted in 30.0 ± 0.5 s, what is the heart rate and its uncertainty in beats per minute?

Solution:

80 ± 3 beats/min

Exercise:

Problem: What is the area of a circle 3.102 cm in diameter?

Exercise:

Problem:

If a marathon runner averages 9.5 mi/h, how long does it take him or her to run a 26.22-mi marathon?

Solution:

2.8 h

Exercise:

Problem:

A marathon runner completes a 42.188-km course in 2 h, 30 min, and 12 s. There is an uncertainty of 25 m in the distance traveled and an uncertainty of 1 s in the elapsed time. (a) Calculate the percent uncertainty in the distance. (b) Calculate the uncertainty in the elapsed time. (c) What is the average speed in meters per second? (d) What is the uncertainty in the average speed?

Exercise:

Problem:

The sides of a small rectangular box are measured to be 1.80 ± 0.01 cm, 2.05 ± 0.02 cm, and 3.1 ± 0.1 cm long. Calculate its volume and uncertainty in cubic centimeters.

Solution:

11 ± 1 cm³

Exercise:

Problem:

When non-metric units were used in the United Kingdom, a unit of mass called the *pound-mass* (lbm) was employed, where $1 \text{ lbm} = 0.4539 \text{ kg}$. (a) If there is an uncertainty of 0.0001 kg in the pound-mass unit, what is its percent uncertainty? (b) Based on that percent uncertainty, what mass in pound-mass has an uncertainty of 1 kg when converted to kilograms?

Exercise:**Problem:**

The length and width of a rectangular room are measured to be $3.955 \pm 0.005 \text{ m}$ and $3.050 \pm 0.005 \text{ m}$. Calculate the area of the room and its uncertainty in square meters.

Solution:

$$12.06 \pm 0.04 \text{ m}^2$$

Exercise:**Problem:**

A car engine moves a piston with a circular cross section of $7.500 \pm 0.002 \text{ cm}$ diameter a distance of $3.250 \pm 0.001 \text{ cm}$ to compress the gas in the cylinder. (a) By what amount is the gas decreased in volume in cubic centimeters? (b) Find the uncertainty in this volume.

Glossary

accuracy

the degree to which a measured value agrees with correct value for that measurement

method of adding percents

the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation

percent uncertainty

the ratio of the uncertainty of a measurement to the measured value, expressed as a percentage

precision

the degree to which repeated measurements agree with each other

significant figures

express the precision of a measuring tool used to measure a value

uncertainty

a quantitative measure of how much your measured values deviate from a standard or expected value

Approximation

- Make reasonable approximations based on given data.

On many occasions, physicists, other scientists, and engineers need to make **approximations** or “guesstimates” for a particular quantity. What is the distance to a certain destination? What is the approximate density of a given item? About how large a current will there be in a circuit? Many approximate numbers are based on formulae in which the input quantities are known only to a limited accuracy. As you develop problem-solving skills (that can be applied to a variety of fields through a study of physics), you will also develop skills at approximating. You will develop these skills through thinking more quantitatively, and by being willing to take risks. As with any endeavor, experience helps, as well as familiarity with units. These approximations allow us to rule out certain scenarios or unrealistic numbers. Approximations also allow us to challenge others and guide us in our approaches to our scientific world. Let us do two examples to illustrate this concept.

Example:

Approximate the Height of a Building

Can you approximate the height of one of the buildings on your campus, or in your neighborhood? Let us make an approximation based upon the height of a person. In this example, we will calculate the height of a 39-story building.

Strategy

Think about the average height of an adult male. We can approximate the height of the building by scaling up from the height of a person.

Solution

Based on information in the example, we know there are 39 stories in the building. If we use the fact that the height of one story is approximately equal to about the length of two adult humans (each human is about 2-m tall), then we can estimate the total height of the building to be

Equation:

$$\frac{2 \text{ m}}{1 \text{ person}} \times \frac{2 \text{ person}}{1 \text{ story}} \times 39 \text{ stories} = 156 \text{ m.}$$

Discussion

You can use known quantities to determine an approximate measurement of unknown quantities. If your hand measures 10 cm across, how many hand lengths equal the width of your desk? What other measurements can you approximate besides length?

Example:

Approximating Vast Numbers: a Trillion Dollars



A bank stack contains one-hundred \$100 bills, and is worth \$10,000. How many bank stacks make up a trillion dollars? (credit: Andrew Magill)

The U.S. federal deficit in the 2008 fiscal year was a little greater than \$10 trillion. Most of us do not have any concept of how much even one trillion actually is. Suppose that you were given a trillion dollars in \$100 bills. If you made 100-bill stacks and used them to evenly cover a football field (between the end zones), make an approximation of how high the money pile would become. (We will use feet/inches rather than meters here)

because football fields are measured in yards.) One of your friends says 3 in., while another says 10 ft. What do you think?

Strategy

When you imagine the situation, you probably envision thousands of small stacks of 100 wrapped \$100 bills, such as you might see in movies or at a bank. Since this is an easy-to-approximate quantity, let us start there. We can find the volume of a stack of 100 bills, find out how many stacks make up one trillion dollars, and then set this volume equal to the area of the football field multiplied by the unknown height.

Solution

(1) Calculate the volume of a stack of 100 bills. The dimensions of a single bill are approximately 3 in. by 6 in. A stack of 100 of these is about 0.5 in. thick. So the total volume of a stack of 100 bills is:

Equation:

$$\begin{aligned}\text{volume of stack} &= \text{length} \times \text{width} \times \text{height}, \\ \text{volume of stack} &= 6 \text{ in.} \times 3 \text{ in.} \times 0.5 \text{ in.}, \\ \text{volume of stack} &= 9 \text{ in.}^3.\end{aligned}$$

(2) Calculate the number of stacks. Note that a trillion dollars is equal to $\$1 \times 10^{12}$, and a stack of one-hundred \$100 bills is equal to \$10,000, or $\$1 \times 10^4$. The number of stacks you will have is:

Equation:

$$\$1 \times 10^{12} (\text{a trillion dollars}) / \$1 \times 10^4 \text{ per stack} = 1 \times 10^8 \text{ stacks.}$$

(3) Calculate the area of a football field in square inches. The area of a football field is 100 yd \times 50 yd, which gives 5,000 yd². Because we are working in inches, we need to convert square yards to square inches:

Equation:

$$\begin{aligned}\text{Area} &= 5,000 \text{ yd}^2 \times \frac{3 \text{ ft}}{1 \text{ yd}} \times \frac{3 \text{ ft}}{1 \text{ yd}} \times \frac{12 \text{ in.}}{1 \text{ ft}} \times \frac{12 \text{ in.}}{1 \text{ ft}} = 6,480,000 \text{ in.}^2, \\ \text{Area} &\approx 6 \times 10^6 \text{ in.}^2.\end{aligned}$$

This conversion gives us $6 \times 10^6 \text{ in.}^2$ for the area of the field. (Note that we are using only one significant figure in these calculations.)

(4) Calculate the total volume of the bills. The volume of all the \$100-bill stacks is $9 \text{ in.}^3/\text{stack} \times 10^8 \text{ stacks} = 9 \times 10^8 \text{ in.}^3$.

(5) Calculate the height. To determine the height of the bills, use the equation:

Equation:

$$\text{volume of bills} = \text{area of field} \times \text{height of money:}$$

$$\text{Height of money} = \frac{\text{volume of bills}}{\text{area of field}},$$

$$\text{Height of money} = \frac{9 \times 10^8 \text{ in.}^3}{6 \times 10^6 \text{ in.}^2} = 1.33 \times 10^2 \text{ in.},$$

$$\text{Height of money} \approx 1 \times 10^2 \text{ in.} = 100 \text{ in.}$$

The height of the money will be about 100 in. high. Converting this value to feet gives

Equation:

$$100 \text{ in.} \times \frac{1 \text{ ft}}{12 \text{ in.}} = 8.33 \text{ ft} \approx 8 \text{ ft.}$$

Discussion

The final approximate value is much higher than the early estimate of 3 in., but the other early estimate of 10 ft (120 in.) was roughly correct. How did the approximation measure up to your first guess? What can this exercise tell you in terms of rough “guesstimates” versus carefully calculated approximations?

Exercise:

Check Your Understanding

Problem:

Using mental math and your understanding of fundamental units, approximate the area of a regulation basketball court. Describe the process you used to arrive at your final approximation.

Solution:

An average male is about two meters tall. It would take approximately 15 men laid out end to end to cover the length, and about 7 to cover the width. That gives an approximate area of 420 m^2 .

Summary

Scientists often approximate the values of quantities to perform calculations and analyze systems.

Problems & Exercises

Exercise:

Problem: How many heartbeats are there in a lifetime?

Solution:

Sample answer: 2×10^9 heartbeats

Exercise:

Problem:

A generation is about one-third of a lifetime. Approximately how many generations have passed since the year 0 AD?

Exercise:

Problem:

How many times longer than the mean life of an extremely unstable atomic nucleus is the lifetime of a human? (Hint: The lifetime of an unstable atomic nucleus is on the order of 10^{-22} s .)

Solution:

Sample answer: 2×10^{31} if an average human lifetime is taken to be about 70 years.

Exercise:

Problem:

Calculate the approximate number of atoms in a bacterium. Assume that the average mass of an atom in the bacterium is ten times the mass of a hydrogen atom. (Hint: The mass of a hydrogen atom is on the order of 10^{-27} kg and the mass of a bacterium is on the order of 10^{-15} kg.)



This color-enhanced photo shows *Salmonella typhimurium* (red) attacking human cells. These bacteria are commonly known for causing foodborne illness. Can you estimate the number of atoms in each bacterium? (credit: Rocky Mountain Laboratories, NIAID, NIH)

Exercise:

Problem:

Approximately how many atoms thick is a cell membrane, assuming all atoms there average about twice the size of a hydrogen atom?

Solution:

Sample answer: 50 atoms

Exercise:**Problem:**

(a) What fraction of Earth's diameter is the greatest ocean depth? (b) The greatest mountain height?

Exercise:**Problem:**

(a) Calculate the number of cells in a hummingbird assuming the mass of an average cell is ten times the mass of a bacterium. (b) Making the same assumption, how many cells are there in a human?

Solution:

Sample answers:

(a) 10^{12} cells/hummingbird

(b) 10^{16} cells/human

Exercise:**Problem:**

Assuming one nerve impulse must end before another can begin, what is the maximum firing rate of a nerve in impulses per second?

Glossary

approximation

an estimated value based on prior experience and reasoning

Introduction to One-Dimensional Kinematics

class="introduction"

The motion
of an
American
kestrel
through the
air can be
described by
the bird's
displacement
, speed,
velocity, and
acceleration.
When it flies
in a straight
line without
any change
in direction,
its motion is
said to be
one
dimensional.
(credit: Vince
Maidens,
Wikimedia
Commons)



Objects are in motion everywhere we look. Everything from a tennis game to a space-probe flyby of the planet Neptune involves motion. When you are resting, your heart moves blood through your veins. And even in inanimate objects, there is continuous motion in the vibrations of atoms and molecules. Questions about motion are interesting in and of themselves: *How long will it take for a space probe to get to Mars? Where will a football land if it is thrown at a certain angle?* But an understanding of motion is also key to understanding other concepts in physics. An understanding of acceleration, for example, is crucial to the study of force.

Our formal study of physics begins with **kinematics** which is defined as the *study of motion without considering its causes*. The word “kinematics” comes from a Greek term meaning motion and is related to other English words such as “cinema” (movies) and “kinesiology” (the study of human motion). In one-dimensional kinematics and [Two-Dimensional Kinematics](#) we will study only the *motion* of a football, for example, without worrying about what forces cause or change its motion. Such considerations come in other chapters. In this chapter, we examine the simplest type of motion—namely, motion along a straight line, or one-dimensional motion. In [Two-Dimensional Kinematics](#), we apply concepts developed here to study motion along curved paths (two- and three-dimensional motion); for example, that of a car rounding a curve.

Displacement

- Define position, displacement, distance, and distance traveled.
- Explain the relationship between position and displacement.
- Distinguish between displacement and distance traveled.
- Calculate displacement and distance given initial position, final position, and the path between the two.



These cyclists in Vietnam can be described by their position relative to buildings and a canal. Their motion can be described by their change in position, or displacement, in the frame of reference. (credit: Suzan Black, Fotopedia)

Position

In order to describe the motion of an object, you must first be able to describe its **position**—where it is at any particular time. More precisely, you need to specify its position relative to a convenient reference frame. Earth is often used as a reference frame, and we often describe the position of an object as it relates to stationary objects in that reference frame. For

example, a rocket launch would be described in terms of the position of the rocket with respect to the Earth as a whole, while a professor's position could be described in terms of where she is in relation to the nearby white board. (See [\[link\]](#).) In other cases, we use reference frames that are not stationary but are in motion relative to the Earth. To describe the position of a person in an airplane, for example, we use the airplane, not the Earth, as the reference frame. (See [\[link\]](#).)

Displacement

If an object moves relative to a reference frame (for example, if a professor moves to the right relative to a white board or a passenger moves toward the rear of an airplane), then the object's position changes. This change in position is known as **displacement**. The word “displacement” implies that an object has moved, or has been displaced.

Note:

Displacement

Displacement is the *change in position* of an object:

Equation:

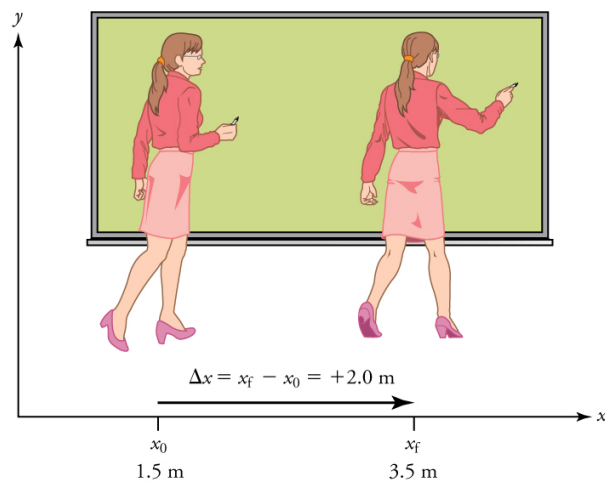
$$\Delta x = x_f - x_0,$$

where Δx is displacement, x_f is the final position, and x_0 is the initial position.

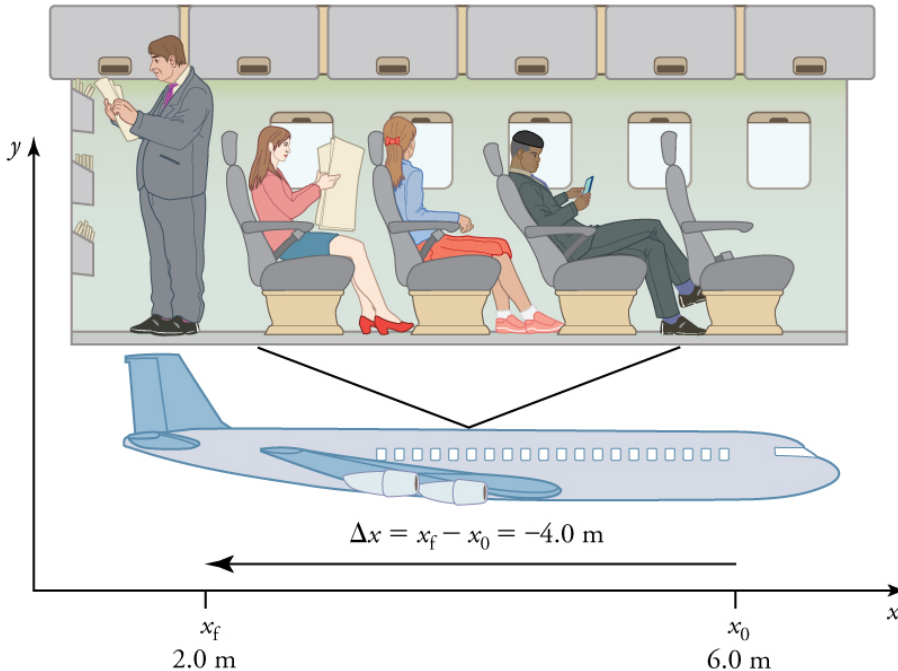
In this text the upper case Greek letter Δ (delta) always means “change in” whatever quantity follows it; thus, Δx means *change in position*. Always solve for displacement by subtracting initial position x_0 from final position x_f .

Note that the SI unit for displacement is the meter (m) (see [Physical Quantities and Units](#)), but sometimes kilometers, miles, feet, and other units of length are used. Keep in mind that when units other than the meter are

used in a problem, you may need to convert them into meters to complete the calculation.



A professor paces left and right while lecturing. Her position relative to Earth is given by x . The $+2.0 \text{ m}$ displacement of the professor relative to Earth is represented by an arrow pointing to the right.



A passenger moves from his seat to the back of the plane. His location relative to the airplane is given by x . The -4.0-m displacement of the passenger relative to the plane is represented by an arrow toward the rear of the plane. Notice that the arrow representing his displacement is twice as long as the arrow representing the displacement of the professor (he moves twice as far) in [\[link\]](#).

Note that displacement has a direction as well as a magnitude. The professor's displacement is 2.0 m to the right, and the airline passenger's displacement is 4.0 m toward the rear. In one-dimensional motion, direction can be specified with a plus or minus sign. When you begin a problem, you should select which direction is positive (usually that will be to the right or up, but you are free to select positive as being any direction). The professor's initial position is $x_0 = 1.5 \text{ m}$ and her final position is $x_f = 3.5 \text{ m}$. Thus her displacement is

Equation:

$$\Delta x = x_f - x_0 = 3.5 \text{ m} - 1.5 \text{ m} = +2.0 \text{ m}.$$

In this coordinate system, motion to the right is positive, whereas motion to the left is negative. Similarly, the airplane passenger's initial position is $x_0 = 6.0$ m and his final position is $x_f = 2.0$ m, so his displacement is

Equation:

$$\Delta x = x_f - x_0 = 2.0 \text{ m} - 6.0 \text{ m} = -4.0 \text{ m}.$$

His displacement is negative because his motion is toward the rear of the plane, or in the negative x direction in our coordinate system.

Distance

Although displacement is described in terms of direction, distance is not. **Distance** is defined to be *the magnitude or size of displacement between two positions*. Note that the distance between two positions is not the same as the distance traveled between them. **Distance traveled** is *the total length of the path traveled between two positions*. Distance has no direction and, thus, no sign. For example, the distance the professor walks is 2.0 m. The distance the airplane passenger walks is 4.0 m.

Note:

Misconception Alert: Distance Traveled vs. Magnitude of Displacement

It is important to note that the *distance traveled*, however, can be greater than the magnitude of the displacement (by magnitude, we mean just the size of the displacement without regard to its direction; that is, just a number with a unit). For example, the professor could pace back and forth many times, perhaps walking a distance of 150 m during a lecture, yet still end up only 2.0 m to the right of her starting point. In this case her displacement would be +2.0 m, the magnitude of her displacement would be 2.0 m, but the distance she traveled would be 150 m. In kinematics we nearly always deal with displacement and magnitude of displacement, and almost never with distance traveled. One way to think about this is to assume you marked the start of the motion and the end of the motion. The

displacement is simply the difference in the position of the two marks and is independent of the path taken in traveling between the two marks. The distance traveled, however, is the total length of the path taken between the two marks.

Exercise:

Check Your Understanding

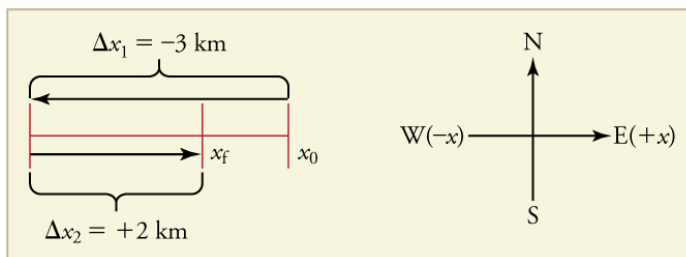
Problem:

A cyclist rides 3 km west and then turns around and rides 2 km east.

(a) What is her displacement? (b) What distance does she ride? (c)

What is the magnitude of her displacement?

Solution:



(a) The rider's displacement is $\Delta x = x_f - x_0 = -1 \text{ km}$. (The displacement is negative because we take east to be positive and west to be negative.)

(b) The distance traveled is $3 \text{ km} + 2 \text{ km} = 5 \text{ km}$.

(c) The magnitude of the displacement is 1 km .

Section Summary

- Kinematics is the study of motion without considering its causes. In this chapter, it is limited to motion along a straight line, called one-dimensional motion.
- Displacement is the change in position of an object.

- In symbols, displacement Δx is defined to be
Equation:

$$\Delta x = x_f - x_0,$$

where x_0 is the initial position and x_f is the final position. In this text, the Greek letter Δ (delta) always means “change in” whatever quantity follows it. The SI unit for displacement is the meter (m). Displacement has a direction as well as a magnitude.

- When you start a problem, assign which direction will be positive.
- Distance is the magnitude of displacement between two positions.
- Distance traveled is the total length of the path traveled between two positions.

Conceptual Questions

Exercise:

Problem:

Give an example in which there are clear distinctions among distance traveled, displacement, and magnitude of displacement. Specifically identify each quantity in your example.

Exercise:

Problem:

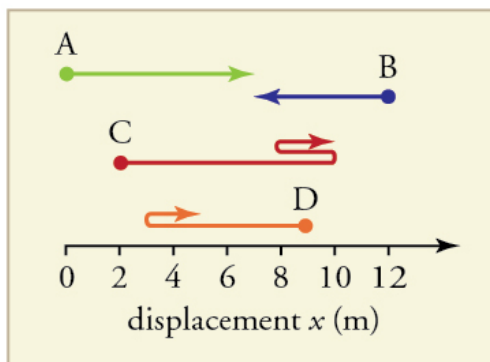
Under what circumstances does distance traveled equal magnitude of displacement? What is the only case in which magnitude of displacement and displacement are exactly the same?

Exercise:

Problem:

Bacteria move back and forth by using their flagella (structures that look like little tails). Speeds of up to $50 \mu\text{m/s}$ ($50 \times 10^{-6} \text{ m/s}$) have been observed. The total distance traveled by a bacterium is large for its size, while its displacement is small. Why is this?

Problems & Exercises



Exercise:

Problem:

Find the following for path A in [\[link\]](#): (a) The distance traveled. (b) The magnitude of the displacement from start to finish. (c) The displacement from start to finish.

Solution:

(a) 7 m

(b) 7 m

(c) +7 m

Exercise:

Problem:

Find the following for path B in [\[link\]](#): (a) The distance traveled. (b) The magnitude of the displacement from start to finish. (c) The displacement from start to finish.

Exercise:

Problem:

Find the following for path C in [\[link\]](#): (a) The distance traveled. (b) The magnitude of the displacement from start to finish. (c) The displacement from start to finish.

Solution:

(a) 13 m

(b) 9 m

(c) +9 m

Exercise:**Problem:**

Find the following for path D in [\[link\]](#): (a) The distance traveled. (b) The magnitude of the displacement from start to finish. (c) The displacement from start to finish.

Glossary

kinematics

the study of motion without considering its causes

position

the location of an object at a particular time

displacement

the change in position of an object

distance

the magnitude of displacement between two positions

distance traveled

the total length of the path traveled between two positions

Vectors, Scalars, and Coordinate Systems

- Define and distinguish between scalar and vector quantities.
- Assign a coordinate system for a scenario involving one-dimensional motion.



The motion of this Eclipse Concept jet can be described in terms of the distance it has traveled (a scalar quantity) or its displacement in a specific direction (a vector quantity). In order to specify the direction of motion, its displacement must be described based on a coordinate system. In this case, it may be convenient to choose motion toward the left as positive motion (it is the forward direction for the plane), although in many cases, the x -coordinate runs from left to right, with motion to the right as positive and motion to the left as negative. (credit: Armchair Aviator, Flickr)

What is the difference between distance and displacement? Whereas displacement is defined by both direction and magnitude, distance is defined only by magnitude. Displacement is an example of a vector quantity. Distance is an example of a scalar quantity. A **vector** is any quantity with both *magnitude and direction*. Other examples of vectors include a velocity of 90 km/h east and a force of 500 newtons straight down.

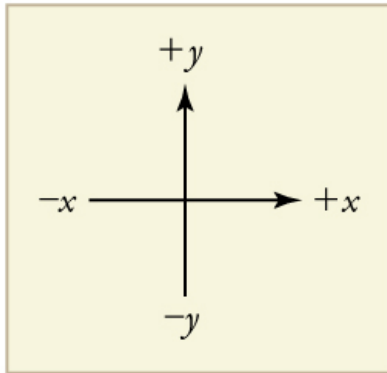
The direction of a vector in one-dimensional motion is given simply by a plus (+) or minus (−) sign. Vectors are represented graphically by arrows. An arrow used to represent a vector has a length proportional to the vector's magnitude (e.g., the larger the magnitude, the longer the length of the vector) and points in the same direction as the vector.

Some physical quantities, like distance, either have no direction or none is specified. A **scalar** is any quantity that has a magnitude, but no direction. For example, a 20°C temperature, the 250 kilocalories (250 Calories) of energy in a candy bar, a 90 km/h speed limit, a person's 1.8 m height, and a distance of 2.0 m are all scalars—quantities with no specified direction. Note, however, that a scalar can be negative, such as a −20°C temperature. In this case, the minus sign indicates a point on a scale rather than a direction. Scalars are never represented by arrows.

Coordinate Systems for One-Dimensional Motion

In order to describe the direction of a vector quantity, you must designate a coordinate system within the reference frame. For one-dimensional motion, this is a simple coordinate system consisting of a one-dimensional coordinate line. In general, when describing horizontal motion, motion to the right is usually considered positive, and motion to the left is considered negative. With vertical motion, motion up is usually positive and motion down is negative. In some cases, however, as with the jet in [\[link\]](#), it can be more convenient to switch the positive and negative directions. For example, if you are analyzing the motion of falling objects, it can be useful to define downwards as the positive direction. If people in a race are

running to the left, it is useful to define left as the positive direction. It does not matter as long as the system is clear and consistent. Once you assign a positive direction and start solving a problem, you cannot change it.



It is usually convenient to consider motion upward or to the right as positive (+) and motion downward or to the left as negative (−).

Exercise:

Check Your Understanding

Problem:

A person's speed can stay the same as he or she rounds a corner and changes direction. Given this information, is speed a scalar or a vector quantity? Explain.

Solution:

Speed is a scalar quantity. It does not change at all with direction changes; therefore, it has magnitude only. If it were a vector quantity, it would change as direction changes (even if its magnitude remained constant).

Section Summary

- A vector is any quantity that has magnitude and direction.
- A scalar is any quantity that has magnitude but no direction.
- Displacement and velocity are vectors, whereas distance and speed are scalars.
- In one-dimensional motion, direction is specified by a plus or minus sign to signify left or right, up or down, and the like.

Conceptual Questions

Exercise:

Problem:

A student writes, “A bird that is diving for prey has a speed of -10 m/s .” What is wrong with the student’s statement? What has the student actually described? Explain.

Exercise:

Problem: What is the speed of the bird in [\[link\]](#)?

Exercise:

Problem:

Acceleration is the change in velocity over time. Given this information, is acceleration a vector or a scalar quantity? Explain.

Exercise:

Problem:

A weather forecast states that the temperature is predicted to be -5°C the following day. Is this temperature a vector or a scalar quantity? Explain.

Glossary

scalar

a quantity that is described by magnitude, but not direction

vector

a quantity that is described by both magnitude and direction

Time, Velocity, and Speed

- Explain the relationships between instantaneous velocity, average velocity, instantaneous speed, average speed, displacement, and time.
- Calculate velocity and speed given initial position, initial time, final position, and final time.
- Derive a graph of velocity vs. time given a graph of position vs. time.
- Interpret a graph of velocity vs. time.



The motion of these racing snails can be described by their speeds and their velocities.
(credit: tobiasflickr, Flickr)

There is more to motion than distance and displacement. Questions such as, “How long does a foot race take?” and “What was the runner’s speed?” cannot be answered without an understanding of other concepts. In this section we add definitions of time, velocity, and speed to expand our description of motion.

Time

As discussed in [Physical Quantities and Units](#), the most fundamental physical quantities are defined by how they are measured. This is the case with time. Every measurement of time involves measuring a change in

some physical quantity. It may be a number on a digital clock, a heartbeat, or the position of the Sun in the sky. In physics, the definition of time is simple—**time** is *change*, or the interval over which change occurs. It is impossible to know that time has passed unless something changes.

The amount of time or change is calibrated by comparison with a standard. The SI unit for time is the second, abbreviated s. We might, for example, observe that a certain pendulum makes one full swing every 0.75 s. We could then use the pendulum to measure time by counting its swings or, of course, by connecting the pendulum to a clock mechanism that registers time on a dial. This allows us to not only measure the amount of time, but also to determine a sequence of events.

How does time relate to motion? We are usually interested in elapsed time for a particular motion, such as how long it takes an airplane passenger to get from his seat to the back of the plane. To find elapsed time, we note the time at the beginning and end of the motion and subtract the two. For example, a lecture may start at 11:00 A.M. and end at 11:50 A.M., so that the elapsed time would be 50 min. **Elapsed time** Δt is the difference between the ending time and beginning time,

Equation:

$$\Delta t = t_f - t_0,$$

where Δt is the change in time or elapsed time, t_f is the time at the end of the motion, and t_0 is the time at the beginning of the motion. (As usual, the delta symbol, Δ , means the change in the quantity that follows it.)

Life is simpler if the beginning time t_0 is taken to be zero, as when we use a stopwatch. If we were using a stopwatch, it would simply read zero at the start of the lecture and 50 min at the end. If $t_0 = 0$, then $\Delta t = t_f \equiv t$.

In this text, for simplicity's sake,

- motion starts at time equal to zero ($t_0 = 0$)
- the symbol t is used for elapsed time unless otherwise specified ($\Delta t = t_f \equiv t$)

Velocity

Your notion of velocity is probably the same as its scientific definition. You know that if you have a large displacement in a small amount of time you have a large velocity, and that velocity has units of distance divided by time, such as miles per hour or kilometers per hour.

Note:

Average Velocity

Average velocity is *displacement (change in position) divided by the time of travel*,

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_0}{t_f - t_0},$$

where \bar{v} is the *average* (indicated by the bar over the v) velocity, Δx is the change in position (or displacement), and x_f and x_0 are the final and beginning positions at times t_f and t_0 , respectively. If the starting time t_0 is taken to be zero, then the average velocity is simply

Equation:

$$\bar{v} = \frac{\Delta x}{t}.$$

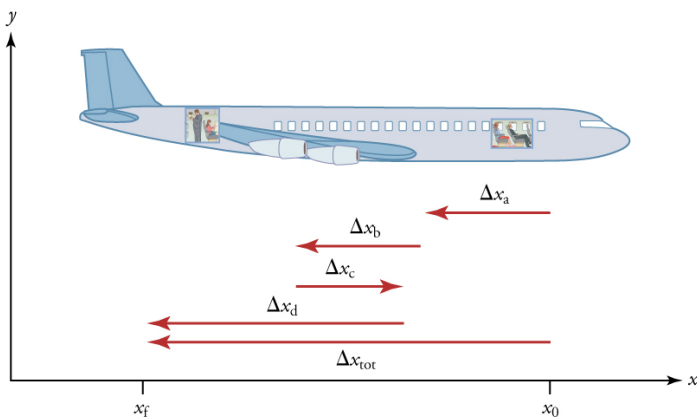
Notice that this definition indicates that *velocity is a vector because displacement is a vector*. It has both magnitude and direction. The SI unit for velocity is meters per second or m/s, but many other units, such as km/h, mi/h (also written as mph), and cm/s, are in common use. Suppose, for example, an airplane passenger took 5 seconds to move -4 m (the negative sign indicates that displacement is toward the back of the plane). His average velocity would be

Equation:

$$\bar{v} = \frac{\Delta x}{t} = \frac{-4 \text{ m}}{5 \text{ s}} = -0.8 \text{ m/s}.$$

The minus sign indicates the average velocity is also toward the rear of the plane.

The average velocity of an object does not tell us anything about what happens to it between the starting point and ending point, however. For example, we cannot tell from average velocity whether the airplane passenger stops momentarily or backs up before he goes to the back of the plane. To get more details, we must consider smaller segments of the trip over smaller time intervals.



A more detailed record of an airplane passenger heading toward the back of the plane, showing smaller segments of his trip.

The smaller the time intervals considered in a motion, the more detailed the information. When we carry this process to its logical conclusion, we are left with an infinitesimally small interval. Over such an interval, the average velocity becomes the *instantaneous velocity* or the *velocity at a specific instant*. A car's speedometer, for example, shows the magnitude (but not the

direction) of the instantaneous velocity of the car. (Police give tickets based on instantaneous velocity, but when calculating how long it will take to get from one place to another on a road trip, you need to use average velocity.)

Instantaneous velocity v is the average velocity at a specific instant in time (or over an infinitesimally small time interval).

Mathematically, finding instantaneous velocity, v , at a precise instant t can involve taking a limit, a calculus operation beyond the scope of this text.

However, under many circumstances, we can find precise values for instantaneous velocity without calculus.

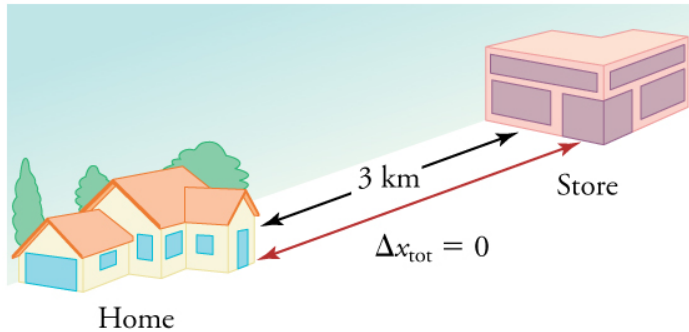
Speed

In everyday language, most people use the terms “speed” and “velocity” interchangeably. In physics, however, they do not have the same meaning and they are distinct concepts. One major difference is that speed has no direction. Thus *speed is a scalar*. Just as we need to distinguish between instantaneous velocity and average velocity, we also need to distinguish between instantaneous speed and average speed.

Instantaneous speed is the magnitude of instantaneous velocity. For example, suppose the airplane passenger at one instant had an instantaneous velocity of -3.0 m/s (the minus meaning toward the rear of the plane). At that same time his instantaneous speed was 3.0 m/s. Or suppose that at one time during a shopping trip your instantaneous velocity is 40 km/h due north. Your instantaneous speed at that instant would be 40 km/h—the same magnitude but without a direction. Average speed, however, is very different from average velocity. **Average speed** is the distance traveled divided by elapsed time.

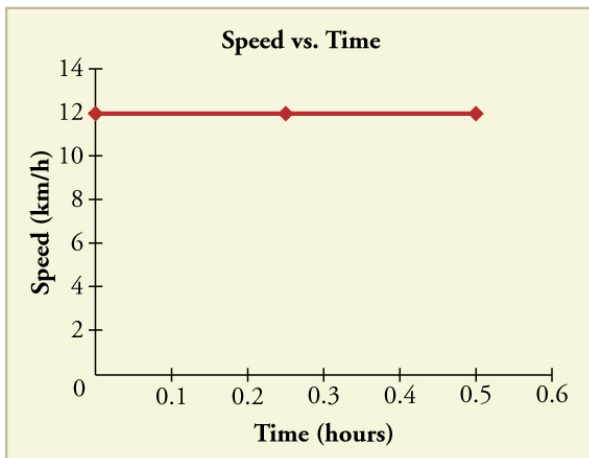
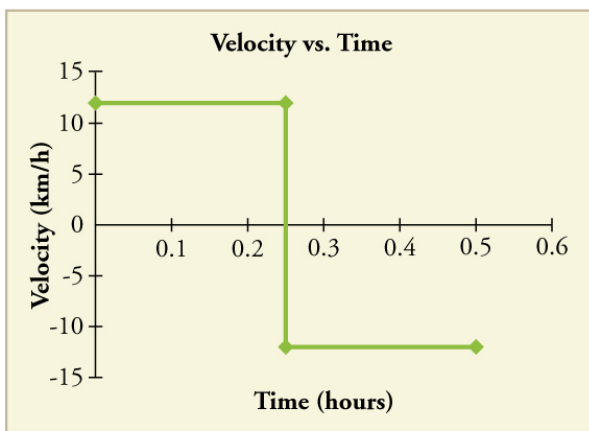
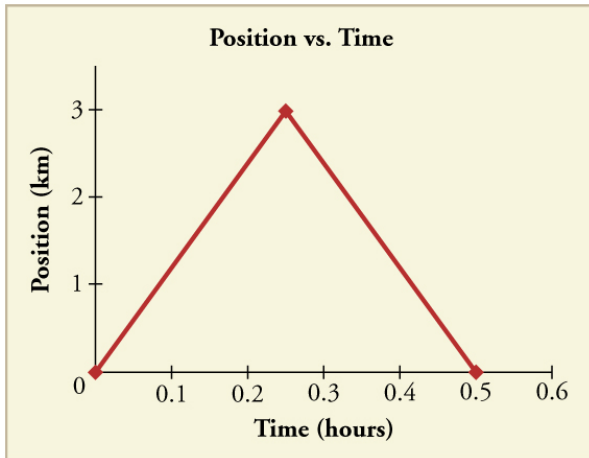
We have noted that distance traveled can be greater than displacement. So average speed can be greater than average velocity, which is displacement divided by time. For example, if you drive to a store and return home in half an hour, and your car’s odometer shows the total distance traveled was 6 km, then your average speed was 12 km/h. Your average velocity, however, was zero, because your displacement for the round trip is zero.

(Displacement is change in position and, thus, is zero for a round trip.) Thus average speed is *not* simply the magnitude of average velocity.



During a 30-minute round trip to the store, the total distance traveled is 6 km. The average speed is 12 km/h. The displacement for the round trip is zero, since there was no net change in position. Thus the average velocity is zero.

Another way of visualizing the motion of an object is to use a graph. A plot of position or of velocity as a function of time can be very useful. For example, for this trip to the store, the position, velocity, and speed-vs.-time graphs are displayed in [\[link\]](#). (Note that these graphs depict a very simplified **model** of the trip. We are assuming that speed is constant during the trip, which is unrealistic given that we'll probably stop at the store. But for simplicity's sake, we will model it with no stops or changes in speed. We are also assuming that the route between the store and the house is a perfectly straight line.)



Position vs. time, velocity vs. time, and speed vs. time on a trip. Note that the velocity for the return trip is negative.

Note:**Making Connections: Take-Home Investigation—Getting a Sense of Speed**

If you have spent much time driving, you probably have a good sense of speeds between about 10 and 70 miles per hour. But what are these in meters per second? What do we mean when we say that something is moving at 10 m/s? To get a better sense of what these values really mean, do some observations and calculations on your own:

- calculate typical car speeds in meters per second
- estimate jogging and walking speed by timing yourself; convert the measurements into both m/s and mi/h
- determine the speed of an ant, snail, or falling leaf

Exercise:**Check Your Understanding****Problem:**

A commuter train travels from Baltimore to Washington, DC, and back in 1 hour and 45 minutes. The distance between the two stations is approximately 40 miles. What is (a) the average velocity of the train, and (b) the average speed of the train in m/s?

Solution:

(a) The average velocity of the train is zero because $x_f = x_0$; the train ends up at the same place it starts.

(b) The average speed of the train is calculated below. Note that the train travels 40 miles one way and 40 miles back, for a total distance of 80 miles.

Equation:

$$\frac{\text{distance}}{\text{time}} = \frac{80 \text{ miles}}{105 \text{ minutes}}$$

Equation:

$$\frac{80 \text{ miles}}{105 \text{ minutes}} \times \frac{5280 \text{ feet}}{1 \text{ mile}} \times \frac{1 \text{ meter}}{3.28 \text{ feet}} \times \frac{1 \text{ minute}}{60 \text{ seconds}} = 20 \text{ m/s}$$

Section Summary

- Time is measured in terms of change, and its SI unit is the second (s). Elapsed time for an event is

Equation:

$$\Delta t = t_f - t_0,$$

where t_f is the final time and t_0 is the initial time. The initial time is often taken to be zero, as if measured with a stopwatch; the elapsed time is then just t .

- Average velocity \bar{v} is defined as displacement divided by the travel time. In symbols, average velocity is

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_0}{t_f - t_0}.$$

- The SI unit for velocity is m/s.
- Velocity is a vector and thus has a direction.
- Instantaneous velocity v is the velocity at a specific instant or the average velocity for an infinitesimal interval.
- Instantaneous speed is the magnitude of the instantaneous velocity.
- Instantaneous speed is a scalar quantity, as it has no direction specified.
- Average speed is the total distance traveled divided by the elapsed time. (Average speed is *not* the magnitude of the average velocity.) Speed is a scalar quantity; it has no direction associated with it.

Conceptual Questions

Exercise:**Problem:**

Give an example (but not one from the text) of a device used to measure time and identify what change in that device indicates a change in time.

Exercise:**Problem:**

There is a distinction between average speed and the magnitude of average velocity. Give an example that illustrates the difference between these two quantities.

Exercise:**Problem:**

Does a car's odometer measure position or displacement? Does its speedometer measure speed or velocity?

Exercise:**Problem:**

If you divide the total distance traveled on a car trip (as determined by the odometer) by the time for the trip, are you calculating the average speed or the magnitude of the average velocity? Under what circumstances are these two quantities the same?

Exercise:**Problem:**

How are instantaneous velocity and instantaneous speed related to one another? How do they differ?

Problems & Exercises**Exercise:**

Problem:

(a) Calculate Earth's average speed relative to the Sun. (b) What is its average velocity over a period of one year?

Solution:

(a) $3.0 \times 10^4 \text{ m/s}$

(b) 0 m/s

Exercise:**Problem:**

A helicopter blade spins at exactly 100 revolutions per minute. Its tip is 5.00 m from the center of rotation. (a) Calculate the average speed of the blade tip in the helicopter's frame of reference. (b) What is its average velocity over one revolution?

Exercise:**Problem:**

The North American and European continents are moving apart at a rate of about 3 cm/y. At this rate how long will it take them to drift 500 km farther apart than they are at present?

Solution:

2×10^7 years

Exercise:

Problem:

Land west of the San Andreas fault in southern California is moving at an average velocity of about 6 cm/y northwest relative to land east of the fault. Los Angeles is west of the fault and may thus someday be at the same latitude as San Francisco, which is east of the fault. How far in the future will this occur if the displacement to be made is 590 km northwest, assuming the motion remains constant?

Exercise:**Problem:**

On May 26, 1934, a streamlined, stainless steel diesel train called the Zephyr set the world's nonstop long-distance speed record for trains. Its run from Denver to Chicago took 13 hours, 4 minutes, 58 seconds, and was witnessed by more than a million people along the route. The total distance traveled was 1633.8 km. What was its average speed in km/h and m/s?

Solution:

$$34.689 \text{ m/s} = 124.88 \text{ km/h}$$

Exercise:**Problem:**

Tidal friction is slowing the rotation of the Earth. As a result, the orbit of the Moon is increasing in radius at a rate of approximately 4 cm/year. Assuming this to be a constant rate, how many years will pass before the radius of the Moon's orbit increases by $3.84 \times 10^6 \text{ m}$ (1%)?

Exercise:

Problem:

A student drove to the university from her home and noted that the odometer reading of her car increased by 12.0 km. The trip took 18.0 min. (a) What was her average speed? (b) If the straight-line distance from her home to the university is 10.3 km in a direction 25.0° south of east, what was her average velocity? (c) If she returned home by the same path 7 h 30 min after she left, what were her average speed and velocity for the entire trip?

Solution:

(a) 40.0 km/h

(b) 34.3 km/h, 25° S of E.

(c) average speed = 3.20 km/h, $\bar{v} = 0$.

Exercise:**Problem:**

The speed of propagation of the action potential (an electrical signal) in a nerve cell depends (inversely) on the diameter of the axon (nerve fiber). If the nerve cell connecting the spinal cord to your feet is 1.1 m long, and the nerve impulse speed is 18 m/s, how long does it take for the nerve signal to travel this distance?

Exercise:

Problem:

Conversations with astronauts on the lunar surface were characterized by a kind of echo in which the earthbound person's voice was so loud in the astronaut's space helmet that it was picked up by the astronaut's microphone and transmitted back to Earth. It is reasonable to assume that the echo time equals the time necessary for the radio wave to travel from the Earth to the Moon and back (that is, neglecting any time delays in the electronic equipment). Calculate the distance from Earth to the Moon given that the echo time was 2.56 s and that radio waves travel at the speed of light (3.00×10^8 m/s).

Solution:

384,000 km

Exercise:**Problem:**

A football quarterback runs 15.0 m straight down the playing field in 2.50 s. He is then hit and pushed 3.00 m straight backward in 1.75 s. He breaks the tackle and runs straight forward another 21.0 m in 5.20 s. Calculate his average velocity (a) for each of the three intervals and (b) for the entire motion.

Exercise:**Problem:**

The planetary model of the atom pictures electrons orbiting the atomic nucleus much as planets orbit the Sun. In this model you can view hydrogen, the simplest atom, as having a single electron in a circular orbit 1.06×10^{-10} m in diameter. (a) If the average speed of the electron in this orbit is known to be 2.20×10^6 m/s, calculate the number of revolutions per second it makes about the nucleus. (b) What is the electron's average velocity?

Solution:

(a) $6.61 \times 10^{15} \text{ rev/s}$

(b) 0 m/s

Glossary

average speed

distance traveled divided by time during which motion occurs

average velocity

displacement divided by time over which displacement occurs

instantaneous velocity

velocity at a specific instant, or the average velocity over an infinitesimal time interval

instantaneous speed

magnitude of the instantaneous velocity

time

change, or the interval over which change occurs

model

simplified description that contains only those elements necessary to describe the physics of a physical situation

elapsed time

the difference between the ending time and beginning time

Acceleration

- Define and distinguish between instantaneous acceleration, average acceleration, and deceleration.
- Calculate acceleration given initial time, initial velocity, final time, and final velocity.



A plane decelerates, or slows down, as it comes in for landing in St. Maarten. Its acceleration is opposite in direction to its velocity. (credit: Steve Conry, Flickr)

In everyday conversation, to accelerate means to speed up. The accelerator in a car can in fact cause it to speed up. The greater the **acceleration**, the greater the change in velocity over a given time. The formal definition of acceleration is consistent with these notions, but more inclusive.

Note:

Average Acceleration

Average Acceleration is *the rate at which velocity changes*,

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0},$$

where \bar{a} is average acceleration, v is velocity, and t is time. (The bar over the a means *average* acceleration.)

Because acceleration is velocity in m/s divided by time in s, the SI units for acceleration are m/s^2 , meters per second squared or meters per second per second, which literally means by how many meters per second the velocity changes every second.

Recall that velocity is a vector—it has both magnitude and direction. This means that a change in velocity can be a change in magnitude (or speed), but it can also be a change in *direction*. For example, if a car turns a corner at constant speed, it is accelerating because its direction is changing. The quicker you turn, the greater the acceleration. So there is an acceleration when velocity changes either in magnitude (an increase or decrease in speed) or in direction, or both.

Note:

Acceleration as a Vector

Acceleration is a vector in the same direction as the *change* in velocity, Δv . Since velocity is a vector, it can change either in magnitude or in direction. Acceleration is therefore a change in either speed or direction, or both.

Keep in mind that although acceleration is in the direction of the *change* in velocity, it is not always in the direction of *motion*. When an object slows down, its acceleration is opposite to the direction of its motion. This is known as **deceleration**.

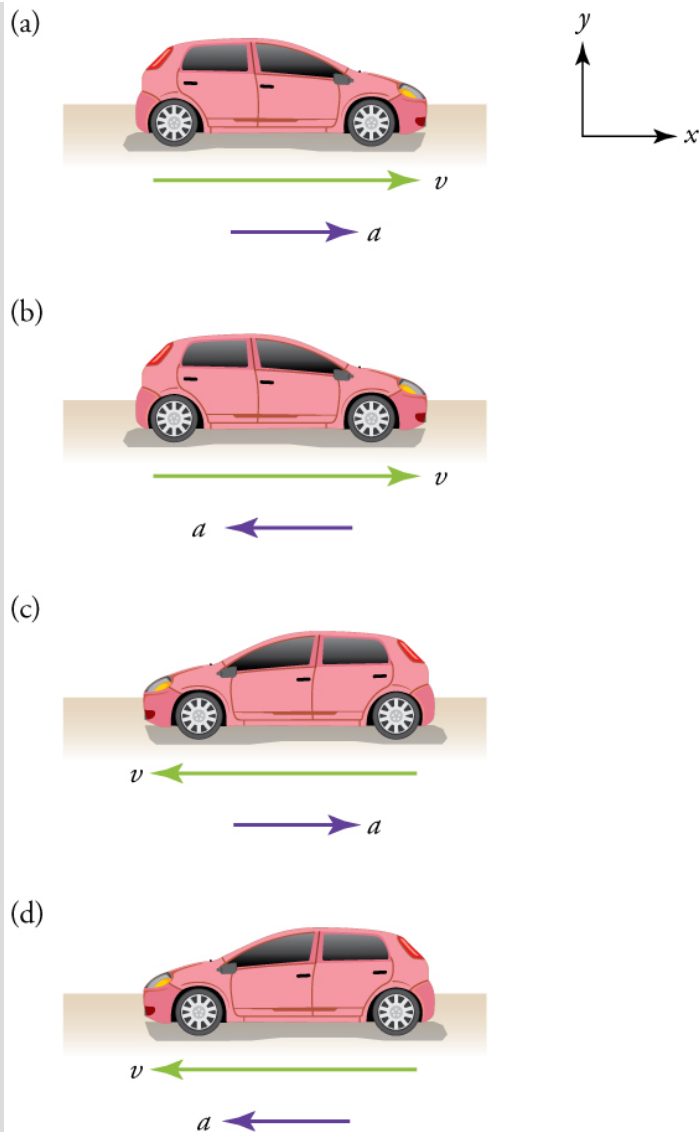


A subway train in Sao Paulo, Brazil, decelerates as it comes into a station. It is accelerating in a direction opposite to its direction of motion. (credit: Yusuke Kawasaki, Flickr)

Note:

Misconception Alert: Deceleration vs. Negative Acceleration

Deceleration always refers to acceleration in the direction opposite to the direction of the velocity. Deceleration always reduces speed. Negative acceleration, however, is acceleration *in the negative direction in the chosen coordinate system*. Negative acceleration may or may not be deceleration, and deceleration may or may not be considered negative acceleration. For example, consider [\[link\]](#).



(a) This car is speeding up as it moves toward the right. It therefore has positive acceleration in our coordinate system. (b) This car is slowing down as it moves toward the right. Therefore, it has negative acceleration in our coordinate system, because its acceleration is toward the left. The car is also decelerating: the direction of its acceleration is opposite to its direction of motion. (c) This car is moving

toward the left, but slowing down over time. Therefore, its acceleration is positive in our coordinate system because it is toward the right.

However, the car is decelerating because its acceleration is opposite to its motion. (d) This car is speeding up as it moves toward the left. It has negative acceleration because it is accelerating toward the left. However, because its acceleration is in the same direction as its motion, it is speeding up (*not* decelerating).

Example:

Calculating Acceleration: A Racehorse Leaves the Gate

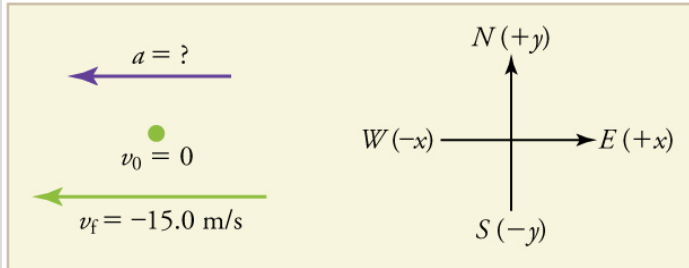
A racehorse coming out of the gate accelerates from rest to a velocity of 15.0 m/s due west in 1.80 s . What is its average acceleration?



(credit: Jon Sullivan, PD
Photo.org)

Strategy

First we draw a sketch and assign a coordinate system to the problem. This is a simple problem, but it always helps to visualize it. Notice that we assign east as positive and west as negative. Thus, in this case, we have negative velocity.



We can solve this problem by identifying Δv and Δt from the given information and then calculating the average acceleration directly from the equation $\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}$.

Solution

1. Identify the knowns. $v_0 = 0$, $v_f = -15.0 \text{ m/s}$ (the negative sign indicates direction toward the west), $\Delta t = 1.80 \text{ s}$.

2. Find the change in velocity. Since the horse is going from zero to -15.0 m/s , its change in velocity equals its final velocity:

$$\Delta v = v_f = -15.0 \text{ m/s}.$$

3. Plug in the known values (Δv and Δt) and solve for the unknown \bar{a} .

Equation:

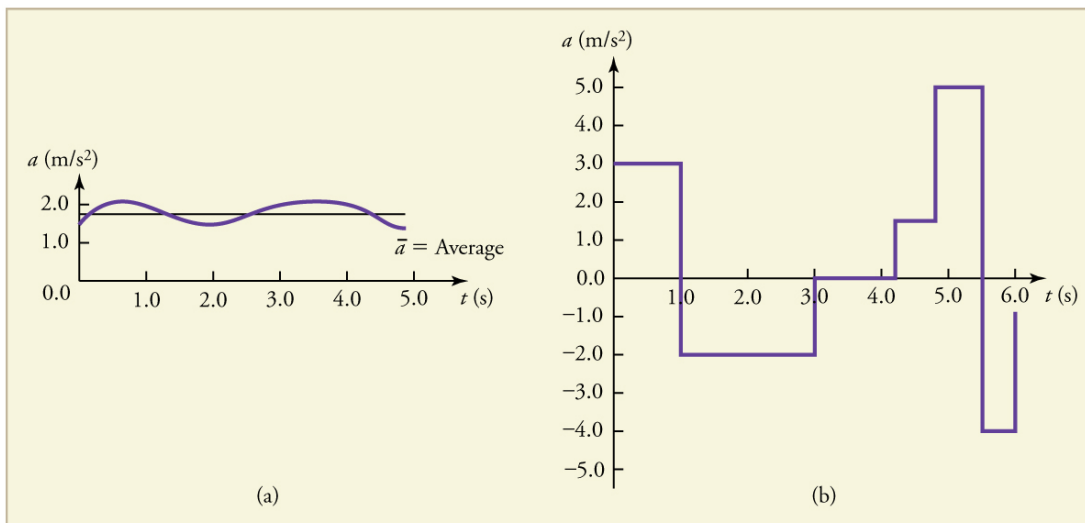
$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-15.0 \text{ m/s}}{1.80 \text{ s}} = -8.33 \text{ m/s}^2.$$

Discussion

The negative sign for acceleration indicates that acceleration is toward the west. An acceleration of 8.33 m/s^2 due west means that the horse increases its velocity by 8.33 m/s due west each second, that is, 8.33 meters per second per second, which we write as 8.33 m/s^2 . This is truly an average acceleration, because the ride is not smooth. We shall see later that an acceleration of this magnitude would require the rider to hang on with a force nearly equal to his weight.

Instantaneous Acceleration

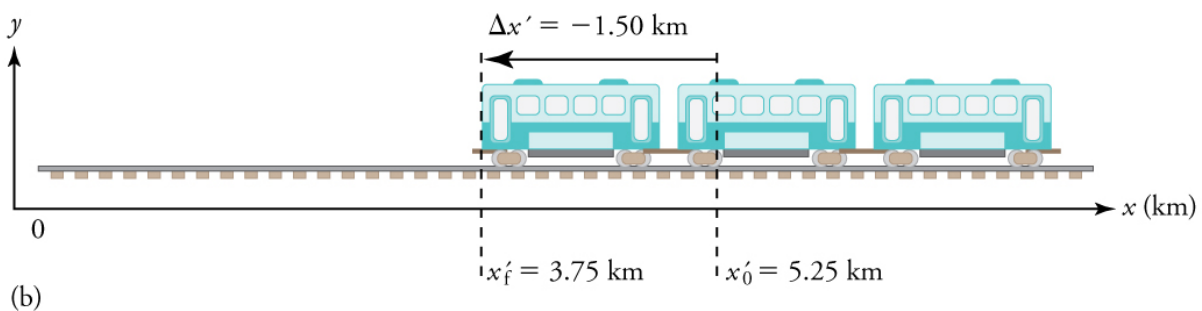
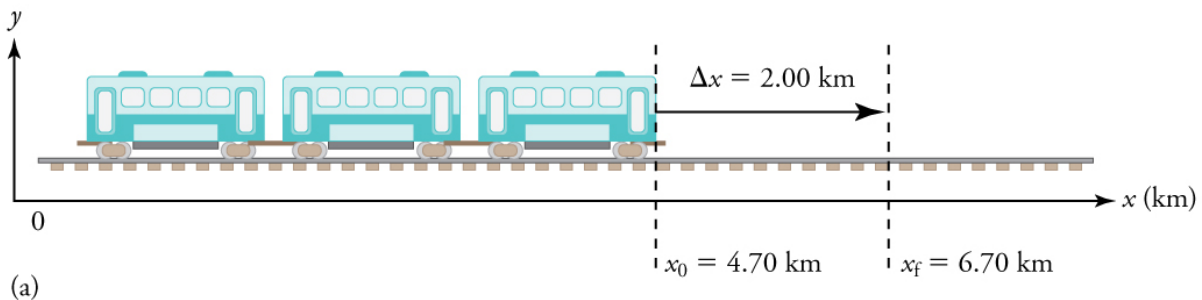
Instantaneous acceleration a , or the *acceleration at a specific instant in time*, is obtained by the same process as discussed for instantaneous velocity in [Time, Velocity, and Speed](#)—that is, by considering an infinitesimally small interval of time. How do we find instantaneous acceleration using only algebra? The answer is that we choose an average acceleration that is representative of the motion. [\[link\]](#) shows graphs of instantaneous acceleration versus time for two very different motions. In [\[link\]](#)(a), the acceleration varies slightly and the average over the entire interval is nearly the same as the instantaneous acceleration at any time. In [\[link\]](#)(b), the acceleration varies drastically over time. In such situations it is best to consider smaller time intervals and choose an average acceleration for each. For example, we could consider motion over the time intervals from 0 to 1.0 s and from 1.0 to 3.0 s as separate motions with accelerations of $+3.0 \text{ m/s}^2$ and -2.0 m/s^2 , respectively.



Graphs of instantaneous acceleration versus time for two different one-dimensional motions. (a) Here acceleration varies only slightly and is always in the same direction, since it is positive. The average over the interval is nearly the same as the

acceleration at any given time. (b) Here the acceleration varies greatly, perhaps representing a package on a post office conveyor belt that is accelerated forward and backward as it bumps along. It is necessary to consider small time intervals (such as from 0 to 1.0 s) with constant or nearly constant acceleration in such a situation.

The next several examples consider the motion of the subway train shown in [\[link\]](#). In (a) the shuttle moves to the right, and in (b) it moves to the left. The examples are designed to further illustrate aspects of motion and to illustrate some of the reasoning that goes into solving problems.



One-dimensional motion of a subway train considered in [\[link\]](#), [\[link\]](#), [\[link\]](#), [\[link\]](#), [\[link\]](#), and [\[link\]](#). Here we have chosen the x -axis so that $+$ means to the right and $-$ means to the left for displacements, velocities, and accelerations. (a) The subway train moves to the right from x_0 to x_f . Its displacement Δx is $+2.0$ km. (b) The train moves to the left from x'_0 to x'_f . Its displacement $\Delta x'$ is

–1.5 km. (Note that the prime symbol (') is used simply to distinguish between displacement in the two different situations. The distances of travel and the size of the cars are on different scales to fit everything into the diagram.)

Example:**Calculating Displacement: A Subway Train**

What are the magnitude and sign of displacements for the motions of the subway train shown in parts (a) and (b) of [\[link\]](#)?

Strategy

A drawing with a coordinate system is already provided, so we don't need to make a sketch, but we should analyze it to make sure we understand what it is showing. Pay particular attention to the coordinate system. To find displacement, we use the equation $\Delta x = x_f - x_0$. This is straightforward since the initial and final positions are given.

Solution

1. Identify the knowns. In the figure we see that $x_f = 6.70$ km and $x_0 = 4.70$ km for part (a), and $x'_f = 3.75$ km and $x'_0 = 5.25$ km for part (b).
2. Solve for displacement in part (a).

Equation:

$$\Delta x = x_f - x_0 = 6.70 \text{ km} - 4.70 \text{ km} = +2.00 \text{ km}$$

3. Solve for displacement in part (b).

Equation:

$$\Delta x' = x'_f - x'_0 = 3.75 \text{ km} - 5.25 \text{ km} = -1.50 \text{ km}$$

Discussion

The direction of the motion in (a) is to the right and therefore its displacement has a positive sign, whereas motion in (b) is to the left and thus has a negative sign.

Example:**Comparing Distance Traveled with Displacement: A Subway Train**

What are the distances traveled for the motions shown in parts (a) and (b) of the subway train in [\[link\]](#)?

Strategy

To answer this question, think about the definitions of distance and distance traveled, and how they are related to displacement. Distance between two positions is defined to be the magnitude of displacement, which was found in [\[link\]](#). Distance traveled is the total length of the path traveled between the two positions. (See [Displacement](#).) In the case of the subway train shown in [\[link\]](#), the distance traveled is the same as the distance between the initial and final positions of the train.

Solution

1. The displacement for part (a) was $+2.00$ km. Therefore, the distance between the initial and final positions was 2.00 km, and the distance traveled was 2.00 km.
2. The displacement for part (b) was -1.5 km. Therefore, the distance between the initial and final positions was 1.50 km, and the distance traveled was 1.50 km.

Discussion

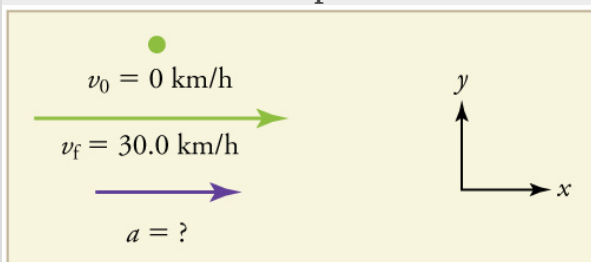
Distance is a scalar. It has magnitude but no sign to indicate direction.

Example:**Calculating Acceleration: A Subway Train Speeding Up**

Suppose the train in [\[link\]](#)(a) accelerates from rest to 30.0 km/h in the first 20.0 s of its motion. What is its average acceleration during that time interval?

Strategy

It is worth it at this point to make a simple sketch:



This problem involves three steps. First we must determine the change in velocity, then we must determine the change in time, and finally we use these values to calculate the acceleration.

Solution

1. Identify the knowns. $v_0 = 0$ (the train starts at rest), $v_f = 30.0 \text{ km/h}$, and $\Delta t = 20.0 \text{ s}$.
2. Calculate Δv . Since the train starts from rest, its change in velocity is $\Delta v = +30.0 \text{ km/h}$, where the plus sign means velocity to the right.
3. Plug in known values and solve for the unknown, \bar{a} .

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{+30.0 \text{ km/h}}{20.0 \text{ s}}$$

4. Since the units are mixed (we have both hours and seconds for time), we need to convert everything into SI units of meters and seconds. (See [Physical Quantities and Units](#) for more guidance.)

Equation:

$$\bar{a} = \left(\frac{+30 \text{ km/h}}{20.0 \text{ s}} \right) \left(\frac{10^3 \text{ m}}{1 \text{ km}} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) = 0.417 \text{ m/s}^2$$

Discussion

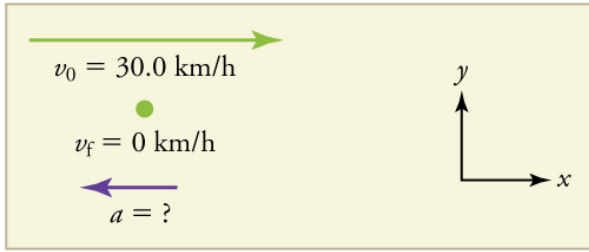
The plus sign means that acceleration is to the right. This is reasonable because the train starts from rest and ends up with a velocity to the right (also positive). So acceleration is in the same direction as the *change* in velocity, as is always the case.

Example:

Calculate Acceleration: A Subway Train Slowing Down

Now suppose that at the end of its trip, the train in [\[link\]](#)(a) slows to a stop from a speed of 30.0 km/h in 8.00 s . What is its average acceleration while stopping?

Strategy



In this case, the train is decelerating and its acceleration is negative because it is toward the left. As in the previous example, we must find the change in velocity and the change in time and then solve for acceleration.

Solution

1. Identify the knowns. $v_0 = 30.0 \text{ km/h}$, $v_f = 0 \text{ km/h}$ (the train is stopped, so its velocity is 0), and $\Delta t = 8.00 \text{ s}$.
2. Solve for the change in velocity, Δv .

Equation:

$$\Delta v = v_f - v_0 = 0 - 30.0 \text{ km/h} = -30.0 \text{ km/h}$$

3. Plug in the knowns, Δv and Δt , and solve for \bar{a} .

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-30.0 \text{ km/h}}{8.00 \text{ s}}$$

4. Convert the units to meters and seconds.

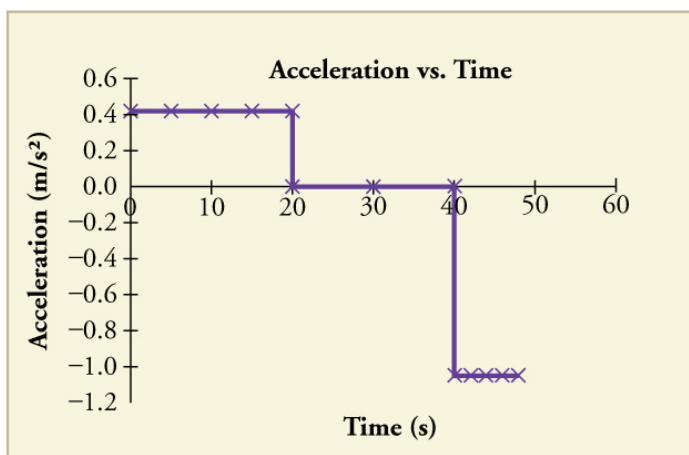
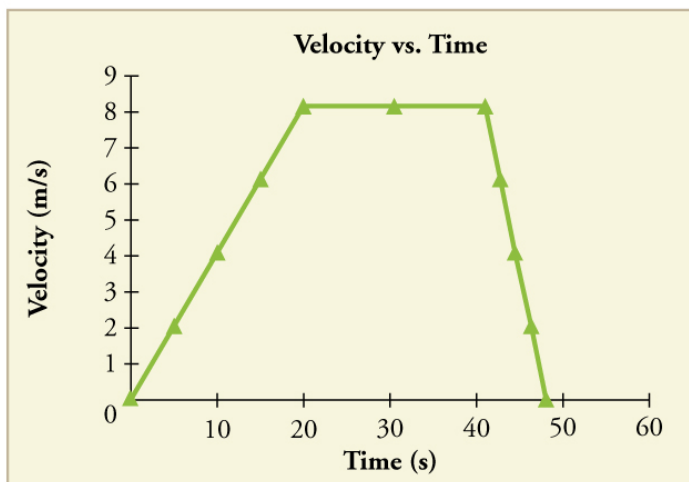
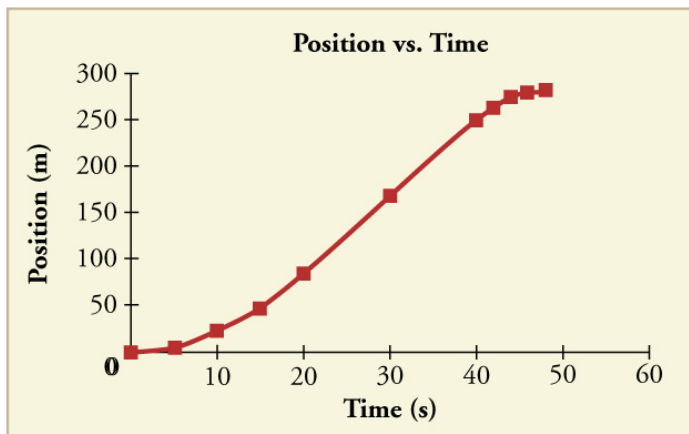
Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \left(\frac{-30.0 \text{ km/h}}{8.00 \text{ s}} \right) \left(\frac{10^3 \text{ m}}{1 \text{ km}} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) = -1.04 \text{ m/s}^2.$$

Discussion

The minus sign indicates that acceleration is to the left. This sign is reasonable because the train initially has a positive velocity in this problem, and a negative acceleration would oppose the motion. Again, acceleration is in the same direction as the *change* in velocity, which is negative here. This acceleration can be called a deceleration because it has a direction opposite to the velocity.

The graphs of position, velocity, and acceleration vs. time for the trains in [\[link\]](#) and [\[link\]](#) are displayed in [\[link\]](#). (We have taken the velocity to remain constant from 20 to 40 s, after which the train decelerates.)



(a) Position of the train over time.

Notice that the train's position changes slowly at the beginning of the journey, then more and more quickly as it picks up speed. Its position then changes more slowly as it slows down at the end of the journey. In the middle of the journey, while the velocity remains constant, the position changes at a constant rate. (b) Velocity

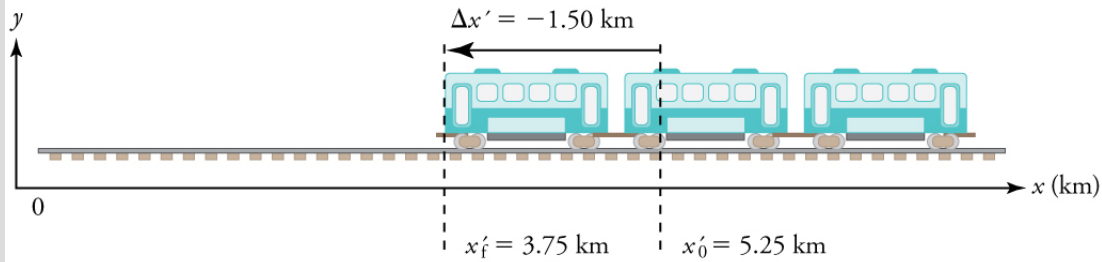
of the train over time. The train's velocity increases as it accelerates at the beginning of the journey. It remains the same in the middle of the journey (where there is no acceleration). It decreases as the train decelerates at the end of the journey.

(c) The acceleration of the train over time. The train has positive acceleration as it speeds up at the beginning of the journey. It has no acceleration as it travels at constant velocity in the middle of the journey. Its acceleration is negative as it slows down at the end of the journey.

Example:

Calculating Average Velocity: The Subway Train

What is the average velocity of the train in part b of [\[link\]](#), and shown again below, if it takes 5.00 min to make its trip?



Strategy

Average velocity is displacement divided by time. It will be negative here, since the train moves to the left and has a negative displacement.

Solution

1. Identify the knowns. $x'_f = 3.75$ km, $x'_0 = 5.25$ km, $\Delta t = 5.00$ min.
2. Determine displacement, $\Delta x'$. We found $\Delta x'$ to be -1.5 km in [\[link\]](#).
3. Solve for average velocity.

Equation:

$$\bar{v} = \frac{\Delta x'}{\Delta t} = \frac{-1.50 \text{ km}}{5.00 \text{ min}}$$

4. Convert units.

Equation:

$$\bar{v} = \frac{\Delta x'}{\Delta t} = \left(\frac{-1.50 \text{ km}}{5.00 \text{ min}} \right) \left(\frac{60 \text{ min}}{1 \text{ h}} \right) = -18.0 \text{ km/h}$$

Discussion

The negative velocity indicates motion to the left.

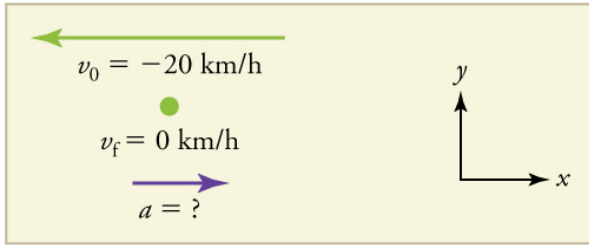
Example:

Calculating Deceleration: The Subway Train

Finally, suppose the train in [\[link\]](#) slows to a stop from a velocity of 20.0 km/h in 10.0 s. What is its average acceleration?

Strategy

Once again, let's draw a sketch:



As before, we must find the change in velocity and the change in time to calculate average acceleration.

Solution

1. Identify the knowns. $v_0 = -20 \text{ km/h}$, $v_f = 0 \text{ km/h}$, $\Delta t = 10.0 \text{ s}$.
2. Calculate Δv . The change in velocity here is actually positive, since

Equation:

$$\Delta v = v_f - v_0 = 0 - (-20 \text{ km/h}) = +20 \text{ km/h}.$$

3. Solve for \bar{a} .

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{+20.0 \text{ km/h}}{10.0 \text{ s}}$$

4. Convert units.

Equation:

$$\bar{a} = \left(\frac{+20.0 \text{ km/h}}{10.0 \text{ s}} \right) \left(\frac{10^3 \text{ m}}{1 \text{ km}} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) = +0.556 \text{ m/s}^2$$

Discussion

The plus sign means that acceleration is to the right. This is reasonable because the train initially has a negative velocity (to the left) in this problem and a positive acceleration opposes the motion (and so it is to the right). Again, acceleration is in the same direction as the *change* in velocity, which is positive here. As in [\[link\]](#), this acceleration can be called a deceleration since it is in the direction opposite to the velocity.

Sign and Direction

Perhaps the most important thing to note about these examples is the signs of the answers. In our chosen coordinate system, plus means the quantity is to the right and minus means it is to the left. This is easy to imagine for displacement and velocity. But it is a little less obvious for acceleration. Most people interpret negative acceleration as the slowing of an object. This was not the case in [\[link\]](#), where a positive acceleration slowed a negative velocity. The crucial distinction was that the acceleration was in the opposite direction from the velocity. In fact, a negative acceleration will *increase* a negative velocity. For example, the train moving to the left in [\[link\]](#) is sped up by an acceleration to the left. In that case, both v and a are negative. The plus and minus signs give the directions of the accelerations. If acceleration has the same sign as the velocity, the object is speeding up. If acceleration has the opposite sign as the velocity, the object is slowing down.

Exercise:

Check Your Understanding

Problem:

An airplane lands on a runway traveling east. Describe its acceleration.

Solution:

If we take east to be positive, then the airplane has negative acceleration, as it is accelerating toward the west. It is also decelerating: its acceleration is opposite in direction to its velocity.

Note:

PhET Explorations: Moving Man Simulation

Learn about position, velocity, and acceleration graphs. Move the little man back and forth with the mouse and plot his motion. Set the position, velocity, or acceleration and let the simulation move the man for you.

<https://archive.cnx.org/specials/e2ca52af-8c6b-450e-ac2f-9300b38e8739/moving-man/>

Section Summary

- Acceleration is the rate at which velocity changes. In symbols, **average acceleration** \bar{a} is

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}.$$

- The SI unit for acceleration is m/s^2 .
- Acceleration is a vector, and thus has both a magnitude and direction.
- Acceleration can be caused by either a change in the magnitude or the direction of the velocity.
- Instantaneous acceleration a is the acceleration at a specific instant in time.
- Deceleration is an acceleration with a direction opposite to that of the velocity.

Conceptual Questions

Exercise:

Problem:

Is it possible for speed to be constant while acceleration is not zero?
Give an example of such a situation.

Exercise:

Problem:

Is it possible for velocity to be constant while acceleration is not zero?
Explain.

Exercise:

Problem:

Give an example in which velocity is zero yet acceleration is not.

Exercise:**Problem:**

If a subway train is moving to the left (has a negative velocity) and then comes to a stop, what is the direction of its acceleration? Is the acceleration positive or negative?

Exercise:**Problem:**

Plus and minus signs are used in one-dimensional motion to indicate direction. What is the sign of an acceleration that reduces the magnitude of a negative velocity? Of a positive velocity?

Problems & Exercises**Exercise:****Problem:**

A cheetah can accelerate from rest to a speed of 30.0 m/s in 7.00 s. What is its acceleration?

Solution:

$$4.29 \text{ m/s}^2$$

Exercise:**Problem: Professional Application**

Dr. John Paul Stapp was U.S. Air Force officer who studied the effects of extreme deceleration on the human body. On December 10, 1954, Stapp rode a rocket sled, accelerating from rest to a top speed of 282 m/s (1015 km/h) in 5.00 s, and was brought jarringly back to rest in only 1.40 s! Calculate his (a) acceleration and (b) deceleration.

Express each in multiples of g (9.80 m/s^2) by taking its ratio to the acceleration of gravity.

Exercise:

Problem:

A commuter backs her car out of her garage with an acceleration of 1.40 m/s^2 . (a) How long does it take her to reach a speed of 2.00 m/s ? (b) If she then brakes to a stop in 0.800 s , what is her deceleration?

Solution:

(a) 1.43 s

(b) -2.50 m/s^2

Exercise:

Problem:

Assume that an intercontinental ballistic missile goes from rest to a suborbital speed of 6.50 km/s in 60.0 s (the actual speed and time are classified). What is its average acceleration in m/s^2 and in multiples of g (9.80 m/s^2)?

Glossary

acceleration

the rate of change in velocity; the change in velocity over time

average acceleration

the change in velocity divided by the time over which it changes

instantaneous acceleration

acceleration at a specific point in time

deceleration

acceleration in the direction opposite to velocity; acceleration that results in a decrease in velocity

Motion Equations for Constant Acceleration in One Dimension

- Calculate displacement of an object that is not accelerating, given initial position and velocity.
- Calculate final velocity of an accelerating object, given initial velocity, acceleration, and time.
- Calculate displacement and final position of an accelerating object, given initial position, initial velocity, time, and acceleration.



Kinematic equations can help us describe and predict the motion of moving objects such as these kayaks racing in Newbury, England. (credit: Barry Skeates, Flickr)

We might know that the greater the acceleration of, say, a car moving away from a stop sign, the greater the displacement in a given time. But we have not developed a specific equation that relates acceleration and displacement. In this section, we develop some convenient equations for kinematic relationships, starting from the definitions of displacement, velocity, and acceleration already covered.

Notation: t , x , v , a

First, let us make some simplifications in notation. Taking the initial time to be zero, as if time is measured with a stopwatch, is a great simplification. Since elapsed time is $\Delta t = t_f - t_0$, taking $t_0 = 0$ means that $\Delta t = t_f$, the final time on the stopwatch. When initial time is taken to be zero, we use the subscript 0 to denote initial values of position and velocity. That is, x_0 is *the initial position* and v_0 is *the initial velocity*. We put no subscripts on the final values. That is, t is *the final time*, x is *the final position*, and v is *the final velocity*. This gives a simpler expression for elapsed time—now, $\Delta t = t$. It also simplifies the expression for displacement, which is now $\Delta x = x - x_0$. Also, it simplifies the expression for change in velocity, which is now $\Delta v = v - v_0$. To summarize, using the simplified notation, with the initial time taken to be zero,

Equation:

$$\begin{aligned}\Delta t &= t \\ \Delta x &= x - x_0 \\ \Delta v &= v - v_0\end{aligned}$$

where *the subscript 0 denotes an initial value and the absence of a subscript denotes a final value* in whatever motion is under consideration.

We now make the important assumption that *acceleration is constant*. This assumption allows us to avoid using calculus to find instantaneous acceleration. Since acceleration is constant, the average and instantaneous accelerations are equal. That is,

Equation:

$$\bar{a} = a = \text{constant},$$

so we use the symbol a for acceleration at all times. Assuming acceleration to be constant does not seriously limit the situations we can study nor degrade the accuracy of our treatment. For one thing, acceleration *is* constant in a great number of situations. Furthermore, in many other situations we can accurately describe motion by assuming a constant acceleration equal to the average acceleration for that motion. Finally, in

motions where acceleration changes drastically, such as a car accelerating to top speed and then braking to a stop, the motion can be considered in separate parts, each of which has its own constant acceleration.

Note:

Solving for Displacement (Δx) and Final Position (x) from Average Velocity when Acceleration (a) is Constant

To get our first two new equations, we start with the definition of average velocity:

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t}.$$

Substituting the simplified notation for Δx and Δt yields

Equation:

$$\bar{v} = \frac{x - x_0}{t}.$$

Solving for x yields

Equation:

$$x = x_0 + \bar{v}t,$$

where the average velocity is

Equation:

$$\bar{v} = \frac{v_0 + v}{2} \text{ (constant } a\text{)}.$$

The equation $\bar{v} = \frac{v_0 + v}{2}$ reflects the fact that, when acceleration is constant, \bar{v} is just the simple average of the initial and final velocities. For example, if

you steadily increase your velocity (that is, with constant acceleration) from 30 to 60 km/h, then your average velocity during this steady increase is 45 km/h. Using the equation $\bar{v} = \frac{v_0 + v}{2}$ to check this, we see that

Equation:

$$\bar{v} = \frac{v_0 + v}{2} = \frac{30 \text{ km/h} + 60 \text{ km/h}}{2} = 45 \text{ km/h},$$

which seems logical.

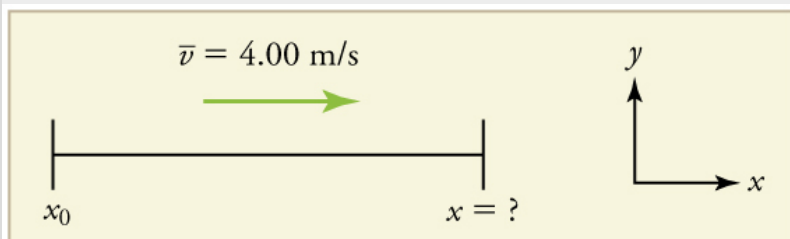
Example:

Calculating Displacement: How Far does the Jogger Run?

A jogger runs down a straight stretch of road with an average velocity of 4.00 m/s for 2.00 min. What is his final position, taking his initial position to be zero?

Strategy

Draw a sketch.



The final position x is given by the equation

Equation:

$$x = x_0 + \bar{v}t.$$

To find x , we identify the values of x_0 , \bar{v} , and t from the statement of the problem and substitute them into the equation.

Solution

1. Identify the knowns. $\bar{v} = 4.00 \text{ m/s}$, $\Delta t = 2.00 \text{ min}$, and $x_0 = 0 \text{ m}$.
2. Enter the known values into the equation.

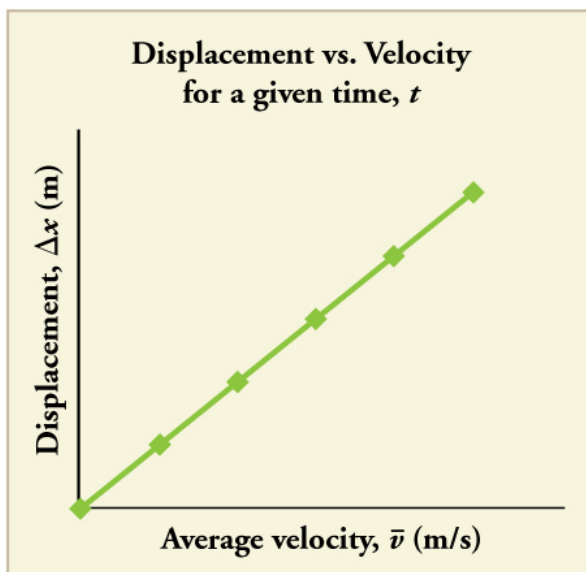
Equation:

$$x = x_0 + \bar{v}t = 0 + (4.00 \text{ m/s})(120 \text{ s}) = 480 \text{ m}$$

Discussion

Velocity and final displacement are both positive, which means they are in the same direction.

The equation $x = x_0 + \bar{v}t$ gives insight into the relationship between displacement, average velocity, and time. It shows, for example, that displacement is a linear function of average velocity. (By linear function, we mean that displacement depends on \bar{v} rather than on \bar{v} raised to some other power, such as \bar{v}^2 . When graphed, linear functions look like straight lines with a constant slope.) On a car trip, for example, we will get twice as far in a given time if we average 90 km/h than if we average 45 km/h.



There is a linear relationship between displacement and average velocity. For a given time t , an object moving twice as fast as another object will

move twice as far as the other object.

Note:

Solving for Final Velocity

We can derive another useful equation by manipulating the definition of acceleration.

Equation:

$$a = \frac{\Delta v}{\Delta t}$$

Substituting the simplified notation for Δv and Δt gives us

Equation:

$$a = \frac{v - v_0}{t} \text{ (constant } a\text{)}.$$

Solving for v yields

Equation:

$$v = v_0 + at \text{ (constant } a\text{)}.$$

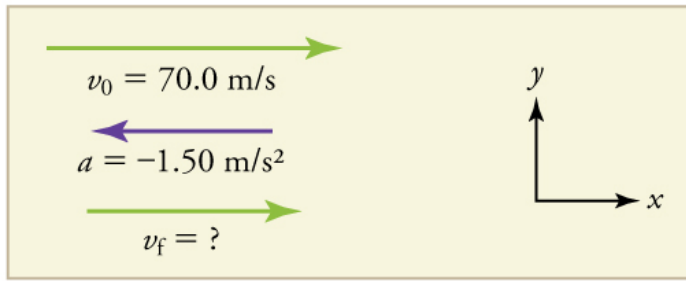
Example:

Calculating Final Velocity: An Airplane Slowing Down after Landing

An airplane lands with an initial velocity of 70.0 m/s and then decelerates at 1.50 m/s² for 40.0 s. What is its final velocity?

Strategy

Draw a sketch. We draw the acceleration vector in the direction opposite the velocity vector because the plane is decelerating.



Solution

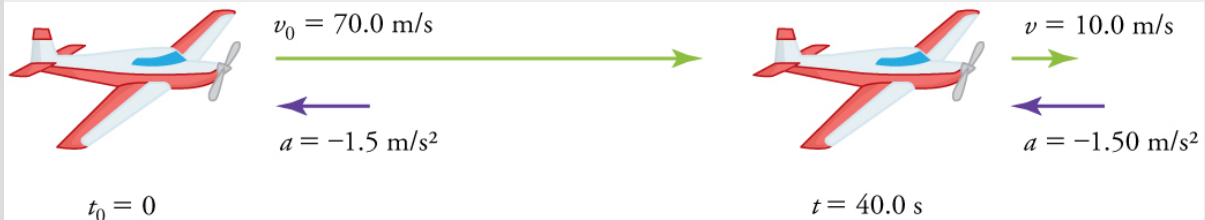
1. Identify the knowns. $v_0 = 70.0 \text{ m/s}$, $a = -1.50 \text{ m/s}^2$, $t = 40.0 \text{ s}$.
2. Identify the unknown. In this case, it is final velocity, v_f .
3. Determine which equation to use. We can calculate the final velocity using the equation $v = v_0 + at$.
4. Plug in the known values and solve.

Equation:

$$v = v_0 + at = 70.0 \text{ m/s} + (-1.50 \text{ m/s}^2)(40.0 \text{ s}) = 10.0 \text{ m/s}$$

Discussion

The final velocity is much less than the initial velocity, as desired when slowing down, but still positive. With jet engines, reverse thrust could be maintained long enough to stop the plane and start moving it backward. That would be indicated by a negative final velocity, which is not the case here.



The airplane lands with an initial velocity of 70.0 m/s and slows to a final velocity of 10.0 m/s before heading for the terminal. Note that the acceleration is negative because its direction is opposite to its velocity, which is positive.

In addition to being useful in problem solving, the equation $v = v_0 + at$ gives us insight into the relationships among velocity, acceleration, and time. From it we can see, for example, that

- final velocity depends on how large the acceleration is and how long it lasts
- if the acceleration is zero, then the final velocity equals the initial velocity ($v = v_0$), as expected (i.e., velocity is constant)
- if a is negative, then the final velocity is less than the initial velocity

(All of these observations fit our intuition, and it is always useful to examine basic equations in light of our intuition and experiences to check that they do indeed describe nature accurately.)

Note:

Making Connections: Real-World Connection



The Space Shuttle *Endeavor*
blasts off from the Kennedy
Space Center in February 2010.
(credit: Matthew Simantov,
Flickr)

An intercontinental ballistic missile (ICBM) has a larger average acceleration than the Space Shuttle and achieves a greater velocity in the

first minute or two of flight (actual ICBM burn times are classified—short-burn-time missiles are more difficult for an enemy to destroy). But the Space Shuttle obtains a greater final velocity, so that it can orbit the earth rather than come directly back down as an ICBM does. The Space Shuttle does this by accelerating for a longer time.

Note:

Solving for Final Position When Velocity is Not Constant ($a \neq 0$)

We can combine the equations above to find a third equation that allows us to calculate the final position of an object experiencing constant acceleration. We start with

Equation:

$$v = v_0 + at.$$

Adding v_0 to each side of this equation and dividing by 2 gives

Equation:

$$\frac{v_0 + v}{2} = v_0 + \frac{1}{2}at.$$

Since $\frac{v_0 + v}{2} = \bar{v}$ for constant acceleration, then

Equation:

$$\bar{v} = v_0 + \frac{1}{2}at.$$

Now we substitute this expression for \bar{v} into the equation for displacement,

$x = x_0 + \bar{v}t$, yielding

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2 \text{ (constant } a\text{)}.$$

Example:**Calculating Displacement of an Accelerating Object: Dragsters**

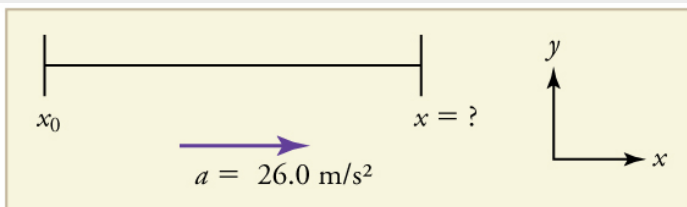
Dragsters can achieve average accelerations of 26.0 m/s^2 . Suppose such a dragster accelerates from rest at this rate for 5.56 s . How far does it travel in this time?



U.S. Army Top Fuel pilot
Tony “The Sarge”
Schumacher begins a race
with a controlled burnout.
(credit: Lt. Col. William
Thurmond. Photo
Courtesy of U.S. Army.)

Strategy

Draw a sketch.



We are asked to find displacement, which is x if we take x_0 to be zero. (Think about it like the starting line of a race. It can be anywhere, but we call it 0 and measure all other positions relative to it.) We can use the equation $x = x_0 + v_0t + \frac{1}{2}at^2$ once we identify v_0 , a , and t from the statement of the problem.

Solution

1. Identify the knowns. Starting from rest means that $v_0 = 0$, a is given as 26.0 m/s^2 and t is given as 5.56 s .
2. Plug the known values into the equation to solve for the unknown x :

Equation:

$$x = x_0 + v_0 t + \frac{1}{2} a t^2.$$

Since the initial position and velocity are both zero, this simplifies to

Equation:

$$x = \frac{1}{2} a t^2.$$

Substituting the identified values of a and t gives

Equation:

$$x = \frac{1}{2} (26.0 \text{ m/s}^2) (5.56 \text{ s})^2,$$

yielding

Equation:

$$x = 402 \text{ m}.$$

Discussion

If we convert 402 m to miles, we find that the distance covered is very close to one quarter of a mile, the standard distance for drag racing. So the answer is reasonable. This is an impressive displacement in only 5.56 s, but top-notch dragsters can do a quarter mile in even less time than this.

What else can we learn by examining the equation $x = x_0 + v_0 t + \frac{1}{2} a t^2$?
We see that:

- displacement depends on the square of the elapsed time when acceleration is not zero. In [\[link\]](#), the dragster covers only one fourth of the total distance in the first half of the elapsed time

- if acceleration is zero, then the initial velocity equals average velocity ($v_0 = \bar{v}$) and $x = x_0 + v_0 t + \frac{1}{2}at^2$ becomes $x = x_0 + v_0 t$

Note:

Solving for Final Velocity when Velocity Is Not Constant ($a \neq 0$)

A fourth useful equation can be obtained from another algebraic manipulation of previous equations.

If we solve $v = v_0 + at$ for t , we get

Equation:

$$t = \frac{v - v_0}{a}.$$

Substituting this and $\bar{v} = \frac{v_0 + v}{2}$ into $x = x_0 + \bar{v}t$, we get

Equation:

$$v^2 = v_0^2 + 2a(x - x_0) \text{ (constant } a\text{)}.$$

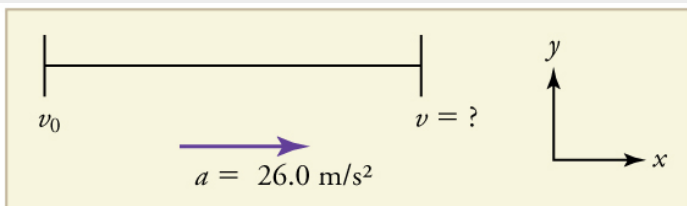
Example:

Calculating Final Velocity: Dragsters

Calculate the final velocity of the dragster in [\[link\]](#) without using information about time.

Strategy

Draw a sketch.



The equation $v^2 = v_0^2 + 2a(x - x_0)$ is ideally suited to this task because it relates velocities, acceleration, and displacement, and no time information is required.

Solution

1. Identify the known values. We know that $v_0 = 0$, since the dragster starts from rest. Then we note that $x - x_0 = 402$ m (this was the answer in [\[link\]](#)). Finally, the average acceleration was given to be $a = 26.0$ m/s².
2. Plug the knowns into the equation $v^2 = v_0^2 + 2a(x - x_0)$ and solve for v .

Equation:

$$v^2 = 0 + 2(26.0 \text{ m/s}^2)(402 \text{ m}).$$

Thus

Equation:

$$v^2 = 2.09 \times 10^4 \text{ m}^2/\text{s}^2.$$

To get v , we take the square root:

Equation:

$$v = \sqrt{2.09 \times 10^4 \text{ m}^2/\text{s}^2} = 145 \text{ m/s}.$$

Discussion

145 m/s is about 522 km/h or about 324 mi/h, but even this breakneck speed is short of the record for the quarter mile. Also, note that a square root has two values; we took the positive value to indicate a velocity in the same direction as the acceleration.

An examination of the equation $v^2 = v_0^2 + 2a(x - x_0)$ can produce further insights into the general relationships among physical quantities:

- The final velocity depends on how large the acceleration is and the distance over which it acts
- For a fixed deceleration, a car that is going twice as fast doesn't simply stop in twice the distance—it takes much further to stop. (This is why

we have reduced speed zones near schools.)

Putting Equations Together

In the following examples, we further explore one-dimensional motion, but in situations requiring slightly more algebraic manipulation. The examples also give insight into problem-solving techniques. The box below provides easy reference to the equations needed.

Note:

Summary of Kinematic Equations (constant a)

Equation:

$$x = x_0 + \bar{v}t$$

Equation:

$$\bar{v} = \frac{v_0 + v}{2}$$

Equation:

$$v = v_0 + at$$

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2$$

Equation:

$$v^2 = v_0^2 + 2a(x - x_0)$$

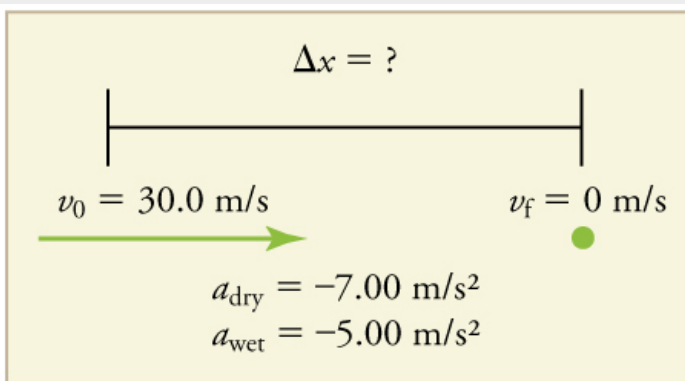
Example:

Calculating Displacement: How Far Does a Car Go When Coming to a Halt?

On dry concrete, a car can decelerate at a rate of 7.00 m/s^2 , whereas on wet concrete it can decelerate at only 5.00 m/s^2 . Find the distances necessary to stop a car moving at 30.0 m/s (about 110 km/h) (a) on dry concrete and (b) on wet concrete. (c) Repeat both calculations, finding the displacement from the point where the driver sees a traffic light turn red, taking into account his reaction time of 0.500 s to get his foot on the brake.

Strategy

Draw a sketch.



In order to determine which equations are best to use, we need to list all of the known values and identify exactly what we need to solve for. We shall do this explicitly in the next several examples, using tables to set them off.

Solution for (a)

1. Identify the knowns and what we want to solve for. We know that $v_0 = 30.0 \text{ m/s}$; $v = 0$; $a = -7.00 \text{ m/s}^2$ (a is negative because it is in a direction opposite to velocity). We take x_0 to be 0. We are looking for displacement Δx , or $x - x_0$.

2. Identify the equation that will help up solve the problem. The best equation to use is

Equation:

$$v^2 = v_0^2 + 2a(x - x_0).$$

This equation is best because it includes only one unknown, x . We know the values of all the other variables in this equation. (There are other equations that would allow us to solve for x , but they require us to know

the stopping time, t , which we do not know. We could use them but it would entail additional calculations.)

3. Rearrange the equation to solve for x .

Equation:

$$x - x_0 = \frac{v^2 - v_0^2}{2a}$$

4. Enter known values.

Equation:

$$x - 0 = \frac{0^2 - (30.0 \text{ m/s})^2}{2(-7.00 \text{ m/s}^2)}$$

Thus,

Equation:

$$x = 64.3 \text{ m on dry concrete.}$$

Solution for (b)

This part can be solved in exactly the same manner as Part A. The only difference is that the deceleration is -5.00 m/s^2 . The result is

Equation:

$$x_{\text{wet}} = 90.0 \text{ m on wet concrete.}$$

Solution for (c)

Once the driver reacts, the stopping distance is the same as it is in Parts A and B for dry and wet concrete. So to answer this question, we need to calculate how far the car travels during the reaction time, and then add that to the stopping time. It is reasonable to assume that the velocity remains constant during the driver's reaction time.

1. Identify the knowns and what we want to solve for. We know that

$\bar{v} = 30.0 \text{ m/s}$; $t_{\text{reaction}} = 0.500 \text{ s}$; $a_{\text{reaction}} = 0$. We take $x_{0-\text{reaction}}$ to be 0. We are looking for x_{reaction} .

2. Identify the best equation to use.

$x = x_0 + \bar{v}t$ works well because the only unknown value is x , which is what we want to solve for.

3. Plug in the knowns to solve the equation.

Equation:

$$x = 0 + (30.0 \text{ m/s})(0.500 \text{ s}) = 15.0 \text{ m}.$$

This means the car travels 15.0 m while the driver reacts, making the total displacements in the two cases of dry and wet concrete 15.0 m greater than if he reacted instantly.

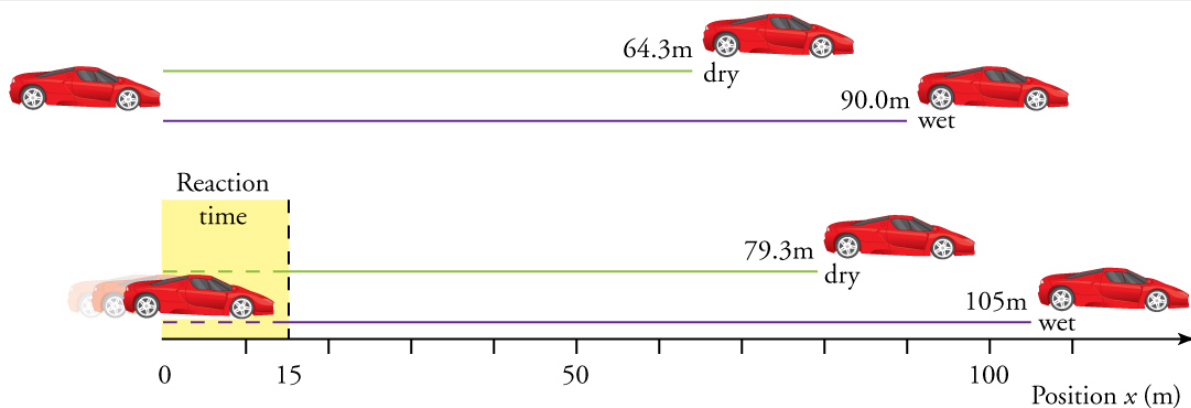
4. Add the displacement during the reaction time to the displacement when braking.

Equation:

$$x_{\text{braking}} + x_{\text{reaction}} = x_{\text{total}}$$

a. $64.3 \text{ m} + 15.0 \text{ m} = 79.3 \text{ m}$ when dry

b. $90.0 \text{ m} + 15.0 \text{ m} = 105 \text{ m}$ when wet



The distance necessary to stop a car varies greatly, depending on road conditions and driver reaction time. Shown here are the braking distances for dry and wet pavement, as calculated in this example, for a car initially traveling at 30.0 m/s. Also shown are the total distances traveled from the point where the driver first sees a light turn red, assuming a 0.500 s reaction time.

Discussion

The displacements found in this example seem reasonable for stopping a fast-moving car. It should take longer to stop a car on wet rather than dry pavement. It is interesting that reaction time adds significantly to the displacements. But more important is the general approach to solving problems. We identify the knowns and the quantities to be determined and then find an appropriate equation. There is often more than one way to solve a problem. The various parts of this example can in fact be solved by other methods, but the solutions presented above are the shortest.

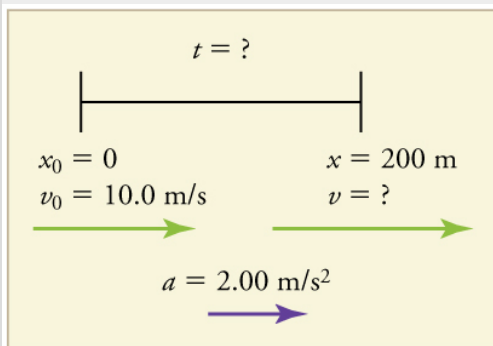
Example:

Calculating Time: A Car Merges into Traffic

Suppose a car merges into freeway traffic on a 200-m-long ramp. If its initial velocity is 10.0 m/s and it accelerates at 2.00 m/s^2 , how long does it take to travel the 200 m up the ramp? (Such information might be useful to a traffic engineer.)

Strategy

Draw a sketch.



We are asked to solve for the time t . As before, we identify the known quantities in order to choose a convenient physical relationship (that is, an equation with one unknown, t).

Solution

1. Identify the knowns and what we want to solve for. We know that $v_0 = 10 \text{ m/s}$; $a = 2.00 \text{ m/s}^2$; and $x = 200 \text{ m}$.
2. We need to solve for t . Choose the best equation. $x = x_0 + v_0 t + \frac{1}{2} a t^2$ works best because the only unknown in the equation is the variable t for which we need to solve.

3. We will need to rearrange the equation to solve for t . In this case, it will be easier to plug in the knowns first.

Equation:

$$200 \text{ m} = 0 \text{ m} + (10.0 \text{ m/s})t + \frac{1}{2} (2.00 \text{ m/s}^2) t^2$$

4. Simplify the equation. The units of meters (m) cancel because they are in each term. We can get the units of seconds (s) to cancel by taking $t = t \text{ s}$, where t is the magnitude of time and s is the unit. Doing so leaves

Equation:

$$200 = 10t + t^2.$$

5. Use the quadratic formula to solve for t .

(a) Rearrange the equation to get 0 on one side of the equation.

Equation:

$$t^2 + 10t - 200 = 0$$

This is a quadratic equation of the form

Equation:

$$at^2 + bt + c = 0,$$

where the constants are $a = 1.00$, $b = 10.0$, and $c = -200$.

(b) Its solutions are given by the quadratic formula:

Equation:

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

This yields two solutions for t , which are

Equation:

$$t = 10.0 \text{ and } -20.0.$$

In this case, then, the time is $t = t$ in seconds, or

Equation:

$$t = 10.0 \text{ s and } -20.0 \text{ s.}$$

A negative value for time is unreasonable, since it would mean that the event happened 20 s before the motion began. We can discard that solution. Thus,

Equation:

$$t = 10.0 \text{ s.}$$

Discussion

Whenever an equation contains an unknown squared, there will be two solutions. In some problems both solutions are meaningful, but in others, such as the above, only one solution is reasonable. The 10.0 s answer seems reasonable for a typical freeway on-ramp.

With the basics of kinematics established, we can go on to many other interesting examples and applications. In the process of developing kinematics, we have also glimpsed a general approach to problem solving that produces both correct answers and insights into physical relationships. [Problem-Solving Basics](#) discusses problem-solving basics and outlines an approach that will help you succeed in this invaluable task.

Note:

Making Connections: Take-Home Experiment—Breaking News

We have been using SI units of meters per second squared to describe some examples of acceleration or deceleration of cars, runners, and trains. To achieve a better feel for these numbers, one can measure the braking deceleration of a car doing a slow (and safe) stop. Recall that, for average acceleration, $\bar{a} = \Delta v / \Delta t$. While traveling in a car, slowly apply the brakes as you come up to a stop sign. Have a passenger note the initial speed in miles per hour and the time taken (in seconds) to stop. From this, calculate the deceleration in miles per hour per second. Convert this to meters per second squared and compare with other decelerations mentioned in this chapter. Calculate the distance traveled in braking.

Exercise:
Check Your Understanding

Problem:

A manned rocket accelerates at a rate of 20 m/s^2 during launch. How long does it take the rocket to reach a velocity of 400 m/s ?

Solution:

To answer this, choose an equation that allows you to solve for time t , given only a , v_0 , and v .

Equation:

$$v = v_0 + at$$

Rearrange to solve for t .

Equation:

$$t = \frac{v - v_0}{a} = \frac{400 \text{ m/s} - 0 \text{ m/s}}{20 \text{ m/s}^2} = 20 \text{ s}$$

Section Summary

- To simplify calculations we take acceleration to be constant, so that $\bar{a} = a$ at all times.
- We also take initial time to be zero.
- Initial position and velocity are given a subscript 0; final values have no subscript. Thus,

Equation:

$$\Delta t = t$$

$$\Delta x = x - x_0$$

$$\Delta v = v - v_0$$

- The following kinematic equations for motion with constant a are useful:

Equation:

$$x = x_0 + \bar{v}t$$

Equation:

$$\bar{v} = \frac{v_0 + v}{2}$$

Equation:

$$v = v_0 + at$$

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2$$

Equation:

$$v^2 = v_0^2 + 2a(x - x_0)$$

- In vertical motion, y is substituted for x .

Problems & Exercises

Exercise:

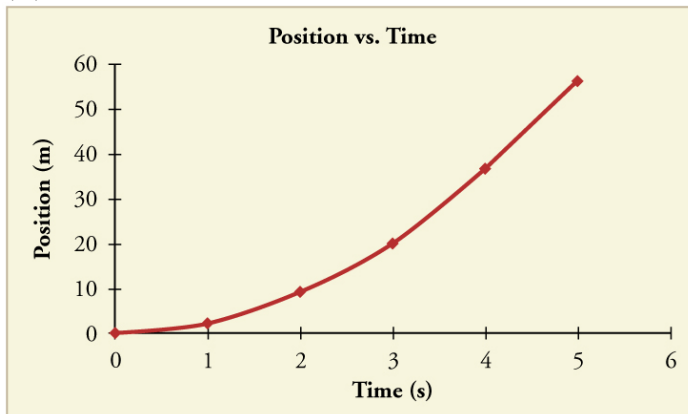
Problem:

An Olympic-class sprinter starts a race with an acceleration of 4.50 m/s^2 . (a) What is her speed 2.40 s later? (b) Sketch a graph of her position vs. time for this period.

Solution:

(a) 10.8 m/s

(b)



Exercise:

Problem:

A well-thrown ball is caught in a well-padded mitt. If the deceleration of the ball is $2.10 \times 10^4 \text{ m/s}^2$, and 1.85 ms ($1 \text{ ms} = 10^{-3} \text{ s}$) elapses from the time the ball first touches the mitt until it stops, what was the initial velocity of the ball?

Solution:

38.9 m/s (about 87 miles per hour)

Exercise:

Problem:

A bullet in a gun is accelerated from the firing chamber to the end of the barrel at an average rate of $6.20 \times 10^5 \text{ m/s}^2$ for $8.10 \times 10^{-4} \text{ s}$. What is its muzzle velocity (that is, its final velocity)?

Exercise:

Problem:

(a) A light-rail commuter train accelerates at a rate of 1.35 m/s^2 . How long does it take to reach its top speed of 80.0 km/h , starting from rest? (b) The same train ordinarily decelerates at a rate of 1.65 m/s^2 . How long does it take to come to a stop from its top speed? (c) In emergencies the train can decelerate more rapidly, coming to rest from 80.0 km/h in 8.30 s . What is its emergency deceleration in m/s^2 ?

Solution:

(a) 16.5 s

(b) 13.5 s

(c) -2.68 m/s^2

Exercise:**Problem:**

While entering a freeway, a car accelerates from rest at a rate of 2.40 m/s^2 for 12.0 s . (a) Draw a sketch of the situation. (b) List the knowns in this problem. (c) How far does the car travel in those 12.0 s ? To solve this part, first identify the unknown, and then discuss how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, check your units, and discuss whether the answer is reasonable. (d) What is the car's final velocity? Solve for this unknown in the same manner as in part (c), showing all steps explicitly.

Exercise:**Problem:**

At the end of a race, a runner decelerates from a velocity of 9.00 m/s at a rate of 2.00 m/s^2 . (a) How far does she travel in the next 5.00 s ? (b) What is her final velocity? (c) Evaluate the result. Does it make sense?

Solution:

(a) 20.0 m

(b) -1.00 m/s

(c) This result does not really make sense. If the runner starts at 9.00 m/s and decelerates at 2.00 m/s^2 , then she will have stopped after 4.50 s . If she continues to decelerate, she will be running backwards.

Exercise:**Problem: Professional Application:**

Blood is accelerated from rest to 30.0 cm/s in a distance of 1.80 cm by the left ventricle of the heart. (a) Make a sketch of the situation. (b) List the knowns in this problem. (c) How long does the acceleration take? To solve this part, first identify the unknown, and then discuss how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, checking your units. (d) Is the answer reasonable when compared with the time for a heartbeat?

Exercise:**Problem:**

In a slap shot, a hockey player accelerates the puck from a velocity of 8.00 m/s to 40.0 m/s in the same direction. If this shot takes $3.33 \times 10^{-2} \text{ s}$, calculate the distance over which the puck accelerates.

Solution:

0.799 m

Exercise:

Problem:

A powerful motorcycle can accelerate from rest to 26.8 m/s (100 km/h) in only 3.90 s. (a) What is its average acceleration? (b) How far does it travel in that time?

Exercise:**Problem:**

Freight trains can produce only relatively small accelerations and decelerations. (a) What is the final velocity of a freight train that accelerates at a rate of 0.0500 m/s^2 for 8.00 min, starting with an initial velocity of 4.00 m/s? (b) If the train can slow down at a rate of 0.550 m/s^2 , how long will it take to come to a stop from this velocity? (c) How far will it travel in each case?

Solution:

(a) 28.0 m/s

(b) 50.9 s

(c) 7.68 km to accelerate and 713 m to decelerate

Exercise:**Problem:**

A fireworks shell is accelerated from rest to a velocity of 65.0 m/s over a distance of 0.250 m. (a) How long did the acceleration last? (b) Calculate the acceleration.

Exercise:

Problem:

A swan on a lake gets airborne by flapping its wings and running on top of the water. (a) If the swan must reach a velocity of 6.00 m/s to take off and it accelerates from rest at an average rate of 0.350 m/s^2 , how far will it travel before becoming airborne? (b) How long does this take?

Solution:

(a) 51.4 m

(b) 17.1 s

Exercise:**Problem: Professional Application:**

A woodpecker's brain is specially protected from large decelerations by tendon-like attachments inside the skull. While pecking on a tree, the woodpecker's head comes to a stop from an initial velocity of 0.600 m/s in a distance of only 2.00 mm . (a) Find the acceleration in m/s^2 and in multiples of g ($g = 9.80 \text{ m/s}^2$). (b) Calculate the stopping time. (c) The tendons cradling the brain stretch, making its stopping distance 4.50 mm (greater than the head and, hence, less deceleration of the brain). What is the brain's deceleration, expressed in multiples of g ?

Exercise:**Problem:**

An unwary football player collides with a padded goalpost while running at a velocity of 7.50 m/s and comes to a full stop after compressing the padding and his body 0.350 m . (a) What is his deceleration? (b) How long does the collision last?

Solution:

(a) -80.4 m/s^2

(b) $9.33 \times 10^{-2} \text{ s}$

Exercise:

Problem:

In World War II, there were several reported cases of airmen who jumped from their flaming airplanes with no parachute to escape certain death. Some fell about 20,000 feet (6000 m), and some of them survived, with few life-threatening injuries. For these lucky pilots, the tree branches and snow drifts on the ground allowed their deceleration to be relatively small. If we assume that a pilot's speed upon impact was 123 mph (54 m/s), then what was his deceleration? Assume that the trees and snow stopped him over a distance of 3.0 m.

Exercise:

Problem:

Consider a grey squirrel falling out of a tree to the ground. (a) If we ignore air resistance in this case (only for the sake of this problem), determine a squirrel's velocity just before hitting the ground, assuming it fell from a height of 3.0 m. (b) If the squirrel stops in a distance of 2.0 cm through bending its limbs, compare its deceleration with that of the airman in the previous problem.

Solution:

(a) 7.7 m/s

(b) $-15 \times 10^2 \text{ m/s}^2$. This is about 3 times the deceleration of the pilots, who were falling from thousands of meters high!

Exercise:

Problem:

An express train passes through a station. It enters with an initial velocity of 22.0 m/s and decelerates at a rate of 0.150 m/s^2 as it goes through. The station is 210 m long. (a) How long is the nose of the train in the station? (b) How fast is it going when the nose leaves the station? (c) If the train is 130 m long, when does the end of the train leave the station? (d) What is the velocity of the end of the train as it leaves?

Exercise:**Problem:**

Dragsters can actually reach a top speed of 145 m/s in only 4.45 s —considerably less time than given in [\[link\]](#) and [\[link\]](#). (a) Calculate the average acceleration for such a dragster. (b) Find the final velocity of this dragster starting from rest and accelerating at the rate found in (a) for 402 m (a quarter mile) without using any information on time. (c) Why is the final velocity greater than that used to find the average acceleration? *Hint:* Consider whether the assumption of constant acceleration is valid for a dragster. If not, discuss whether the acceleration would be greater at the beginning or end of the run and what effect that would have on the final velocity.

Solution:

(a) 32.6 m/s^2

(b) 162 m/s

(c) $v > v_{\text{max}}$, because the assumption of constant acceleration is not valid for a dragster. A dragster changes gears, and would have a greater acceleration in first gear than second gear than third gear, etc. The acceleration would be greatest at the beginning, so it would not be accelerating at 32.6 m/s^2 during the last few meters, but substantially less, and the final velocity would be less than 162 m/s .

Exercise:**Problem:**

A bicycle racer sprints at the end of a race to clinch a victory. The racer has an initial velocity of 11.5 m/s and accelerates at the rate of 0.500 m/s^2 for 7.00 s . (a) What is his final velocity? (b) The racer continues at this velocity to the finish line. If he was 300 m from the finish line when he started to accelerate, how much time did he save? (c) One other racer was 5.00 m ahead when the winner started to accelerate, but he was unable to accelerate, and traveled at 11.8 m/s until the finish line. How far ahead of him (in meters and in seconds) did the winner finish?

Exercise:**Problem:**

In 1967, New Zealander Burt Munro set the world record for an Indian motorcycle, on the Bonneville Salt Flats in Utah, with a maximum speed of 183.58 mi/h . The one-way course was 5.00 mi long. Acceleration rates are often described by the time it takes to reach 60.0 mi/h from rest. If this time was 4.00 s , and Burt accelerated at this rate until he reached his maximum speed, how long did it take Burt to complete the course?

Solution:

104 s

Exercise:

Problem:

(a) A world record was set for the men's 100-m dash in the 2008 Olympic Games in Beijing by Usain Bolt of Jamaica. Bolt "coasted" across the finish line with a time of 9.69 s. If we assume that Bolt accelerated for 3.00 s to reach his maximum speed, and maintained that speed for the rest of the race, calculate his maximum speed and his acceleration. (b) During the same Olympics, Bolt also set the world record in the 200-m dash with a time of 19.30 s. Using the same assumptions as for the 100-m dash, what was his maximum speed for this race?

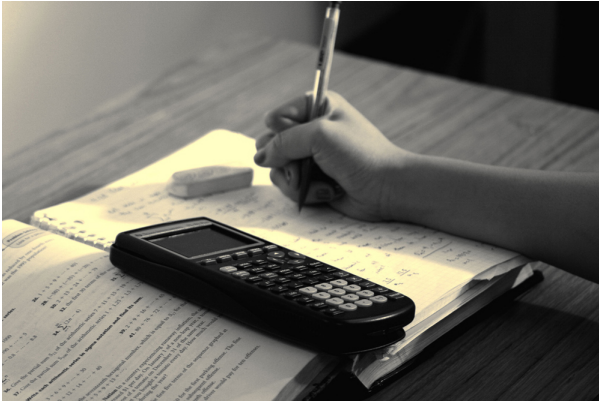
Solution:

(a) $v = 12.2 \text{ m/s}$; $a = 4.07 \text{ m/s}^2$

(b) $v = 11.2 \text{ m/s}$

Problem-Solving Basics for One-Dimensional Kinematics

- Apply problem-solving steps and strategies to solve problems of one-dimensional kinematics.
- Apply strategies to determine whether or not the result of a problem is reasonable, and if not, determine the cause.



Problem-solving skills are essential to your success in Physics. (credit: scui3asteveo, Flickr)

Problem-solving skills are obviously essential to success in a quantitative course in physics. More importantly, the ability to apply broad physical principles, usually represented by equations, to specific situations is a very powerful form of knowledge. It is much more powerful than memorizing a list of facts. Analytical skills and problem-solving abilities can be applied to new situations, whereas a list of facts cannot be made long enough to contain every possible circumstance. Such analytical skills are useful both for solving problems in this text and for applying physics in everyday and professional life.

Problem-Solving Steps

While there is no simple step-by-step method that works for every problem, the following general procedures facilitate problem solving and make it more meaningful. A certain amount of creativity and insight is required as well.

Step 1

Examine the situation to determine which physical principles are involved. It often helps to *draw a simple sketch* at the outset. You will also need to decide which direction is positive and note that on your sketch. Once you have identified the physical principles, it is much easier to find and apply the equations representing those principles. Although finding the correct equation is essential, keep in mind that equations represent physical principles, laws of nature, and relationships among physical quantities. Without a conceptual understanding of a problem, a numerical solution is meaningless.

Step 2

Make a list of what is given or can be inferred from the problem as stated (identify the knowns). Many problems are stated very succinctly and require some inspection to determine what is known. A sketch can also be very useful at this point. Formally identifying the knowns is of particular importance in applying physics to real-world situations. Remember, “stopped” means velocity is zero, and we often can take initial time and position as zero.

Step 3

Identify exactly what needs to be determined in the problem (identify the unknowns). In complex problems, especially, it is not always obvious what needs to be found or in what sequence. Making a list can help.

Step 4

Find an equation or set of equations that can help you solve the problem.

Your list of knowns and unknowns can help here. It is easiest if you can find equations that contain only one unknown—that is, all of the other variables are known, so you can easily solve for the unknown. If the equation contains more than one unknown, then an additional equation is needed to solve the problem. In some problems, several unknowns must be determined to get at the one needed most. In such problems it is especially important to keep physical principles in mind to avoid going astray in a sea of equations. You may have to use two (or more) different equations to get the final answer.

Step 5

Substitute the knowns along with their units into the appropriate equation, and obtain numerical solutions complete with units. This step produces the numerical answer; it also provides a check on units that can help you find errors. If the units of the answer are incorrect, then an error has been made. However, be warned that correct units do not guarantee that the numerical part of the answer is also correct.

Step 6

Check the answer to see if it is reasonable: Does it make sense? This final step is extremely important—the goal of physics is to accurately describe nature. To see if the answer is reasonable, check both its magnitude and its sign, in addition to its units. Your judgment will improve as you solve more and more physics problems, and it will become possible for you to make finer and finer judgments regarding whether nature is adequately described by the answer to a problem. This step brings the problem back to its conceptual meaning. If you can judge whether the answer is reasonable, you have a deeper understanding of physics than just being able to mechanically solve a problem.

When solving problems, we often perform these steps in different order, and we also tend to do several steps simultaneously. There is no rigid procedure that will work every time. Creativity and insight grow with experience, and the basics of problem solving become almost automatic. One way to get practice is to work out the text's examples for yourself as you read. Another is to work as many end-of-section problems as possible, starting with the easiest to build confidence and progressing to the more difficult. Once you become involved in physics, you will see it all around you, and you can begin to apply it to situations you encounter outside the classroom, just as is done in many of the applications in this text.

Unreasonable Results

Physics must describe nature accurately. Some problems have results that are unreasonable because one premise is unreasonable or because certain premises are inconsistent with one another. The physical principle applied correctly then produces an unreasonable result. For example, if a person starting a foot race accelerates at 0.40 m/s^2 for 100 s, his final speed will be 40 m/s (about 150 km/h)—clearly unreasonable because the time of 100 s is an unreasonable premise. The physics is correct in a sense, but there is more to describing nature than just manipulating equations correctly. Checking the result of a problem to see if it is reasonable does more than help uncover errors in problem solving—it also builds intuition in judging whether nature is being accurately described.

Use the following strategies to determine whether an answer is reasonable and, if it is not, to determine what is the cause.

Step 1

Solve the problem using strategies as outlined and in the format followed in the worked examples in the text. In the example given in the preceding paragraph, you would identify the givens as the acceleration and time and use the equation below to find the unknown final velocity. That is,

Equation:

$$v = v_0 + at = 0 + (0.40 \text{ m/s}^2)(100 \text{ s}) = 40 \text{ m/s}.$$

Step 2

Check to see if the answer is reasonable. Is it too large or too small, or does it have the wrong sign, improper units, ...? In this case, you may need to convert meters per second into a more familiar unit, such as miles per hour.

Equation:

$$\left(\frac{40 \text{ m}}{\text{s}}\right) \left(\frac{3.28 \text{ ft}}{\text{m}}\right) \left(\frac{1 \text{ mi}}{5280 \text{ ft}}\right) \left(\frac{60 \text{ s}}{\text{min}}\right) \left(\frac{60 \text{ min}}{1 \text{ h}}\right) = 89 \text{ mph}$$

This velocity is about four times greater than a person can run—so it is too large.

Step 3

If the answer is unreasonable, look for what specifically could cause the identified difficulty. In the example of the runner, there are only two assumptions that are suspect. The acceleration could be too great or the time too long. First look at the acceleration and think about what the number means. If someone accelerates at 0.40 m/s^2 , their velocity is increasing by 0.4 m/s each second. Does this seem reasonable? If so, the time must be too long. It is not possible for someone to accelerate at a constant rate of 0.40 m/s^2 for 100 s (almost two minutes).

Section Summary

- *The six basic problem solving steps for physics are:*

Step 1. Examine the situation to determine which physical principles are involved.

Step 2. Make a list of what is given or can be inferred from the problem as stated (identify the knowns).

Step 3. Identify exactly what needs to be determined in the problem (identify the unknowns).

Step 4. Find an equation or set of equations that can help you solve the problem.

Step 5. Substitute the knowns along with their units into the appropriate equation, and obtain numerical solutions complete with units.

Step 6. Check the answer to see if it is reasonable: Does it make sense?

Conceptual Questions

Exercise:

Problem:

What information do you need in order to choose which equation or equations to use to solve a problem? Explain.

Exercise:

Problem:

What is the last thing you should do when solving a problem? Explain.

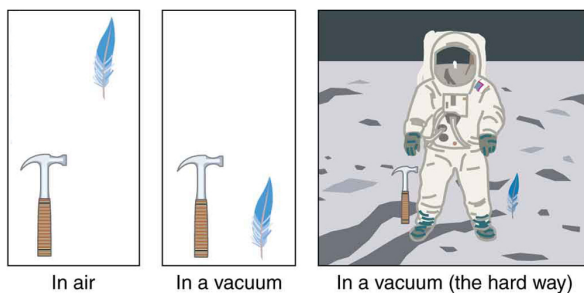
Falling Objects

- Describe the effects of gravity on objects in motion.
- Describe the motion of objects that are in free fall.
- Calculate the position and velocity of objects in free fall.

Falling objects form an interesting class of motion problems. For example, we can estimate the depth of a vertical mine shaft by dropping a rock into it and listening for the rock to hit the bottom. By applying the kinematics developed so far to falling objects, we can examine some interesting situations and learn much about gravity in the process.

Gravity

The most remarkable and unexpected fact about falling objects is that, if air resistance and friction are negligible, then in a given location all objects fall toward the center of Earth with the *same constant acceleration, independent of their mass*. This experimentally determined fact is unexpected, because we are so accustomed to the effects of air resistance and friction that we expect light objects to fall slower than heavy ones.



A hammer and a feather will fall with the same constant acceleration if air resistance is considered negligible. This is a general characteristic of gravity not unique to Earth, as astronaut David R. Scott demonstrated on the Moon in 1971, where the

acceleration due to gravity is
only 1.67 m/s^2 .

In the real world, air resistance can cause a lighter object to fall slower than a heavier object of the same size. A tennis ball will reach the ground after a hard baseball dropped at the same time. (It might be difficult to observe the difference if the height is not large.) Air resistance opposes the motion of an object through the air, while friction between objects—such as between clothes and a laundry chute or between a stone and a pool into which it is dropped—also opposes motion between them. For the ideal situations of these first few chapters, an object *falling without air resistance or friction* is defined to be in **free-fall**.

The force of gravity causes objects to fall toward the center of Earth. The acceleration of free-falling objects is therefore called the **acceleration due to gravity**. The acceleration due to gravity is *constant*, which means we can apply the kinematics equations to any falling object where air resistance and friction are negligible. This opens a broad class of interesting situations to us. The acceleration due to gravity is so important that its magnitude is given its own symbol, g . It is constant at any given location on Earth and has the average value

Equation:

$$g = 9.80 \text{ m/s}^2.$$

Although g varies from 9.78 m/s^2 to 9.83 m/s^2 , depending on latitude, altitude, underlying geological formations, and local topography, the average value of 9.80 m/s^2 will be used in this text unless otherwise specified. The direction of the acceleration due to gravity is *downward (towards the center of Earth)*. In fact, its direction *defines* what we call vertical. Note that whether the acceleration a in the kinematic equations has the value $+g$ or $-g$ depends on how we define our coordinate system. If we define the upward direction as positive, then $a = -g = -9.80 \text{ m/s}^2$, and if we define the downward direction as positive, then $a = g = 9.80 \text{ m/s}^2$.

One-Dimensional Motion Involving Gravity

The best way to see the basic features of motion involving gravity is to start with the simplest situations and then progress toward more complex ones. So we start by considering straight up and down motion with no air resistance or friction. These assumptions mean that the velocity (if there is any) is vertical. If the object is dropped, we know the initial velocity is zero. Once the object has left contact with whatever held or threw it, the object is in free-fall. Under these circumstances, the motion is one-dimensional and has constant acceleration of magnitude g . We will also represent vertical displacement with the symbol y and use x for horizontal displacement.

Note:

Kinematic Equations for Objects in Free-Fall where Acceleration = $-g$

Equation:

$$v = v_0 - gt$$

Equation:

$$y = y_0 + v_0t - \frac{1}{2}gt^2$$

Equation:

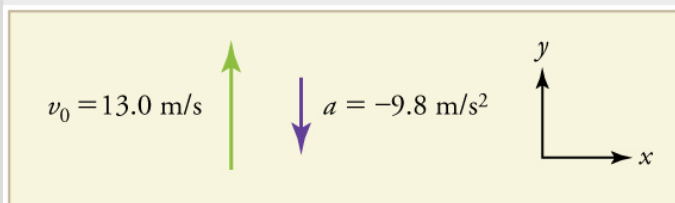
$$v^2 = v_0^2 - 2g(y - y_0)$$

Example:**Calculating Position and Velocity of a Falling Object: A Rock Thrown Upward**

A person standing on the edge of a high cliff throws a rock straight up with an initial velocity of 13.0 m/s. The rock misses the edge of the cliff as it falls back to earth. Calculate the position and velocity of the rock 1.00 s, 2.00 s, and 3.00 s after it is thrown, neglecting the effects of air resistance.

Strategy

Draw a sketch.



We are asked to determine the position y at various times. It is reasonable to take the initial position y_0 to be zero. This problem involves one-dimensional motion in the vertical direction. We use plus and minus signs to indicate direction, with up being positive and down negative. Since up is positive, and the rock is thrown upward, the initial velocity must be positive too. The acceleration due to gravity is downward, so a is negative. It is crucial that the initial velocity and the acceleration due to gravity have opposite signs.

Opposite signs indicate that the acceleration due to gravity opposes the initial motion and will slow and eventually reverse it.

Since we are asked for values of position and velocity at three times, we will refer to these as y_1 and v_1 ; y_2 and v_2 ; and y_3 and v_3 .

Solution for Position y_1

1. Identify the knowns. We know that $y_0 = 0$; $v_0 = 13.0 \text{ m/s}$;

$a = -g = -9.80 \text{ m/s}^2$; and $t = 1.00 \text{ s}$.

2. Identify the best equation to use. We will use $y = y_0 + v_0 t + \frac{1}{2} a t^2$

because it includes only one unknown, y (or y_1 , here), which is the value we want to find.

3. Plug in the known values and solve for y_1 .

Equation:

$$y_1 = 0 + (13.0 \text{ m/s})(1.00 \text{ s}) + \frac{1}{2} (-9.80 \text{ m/s}^2)(1.00 \text{ s})^2 = 8.10 \text{ m}$$

Discussion

The rock is 8.10 m above its starting point at $t = 1.00 \text{ s}$, since $y_1 > y_0$. It could be *moving* up or down; the only way to tell is to calculate v_1 and find out if it is positive or negative.

Solution for Velocity v_1

1. Identify the knowns. We know that $y_0 = 0$; $v_0 = 13.0 \text{ m/s}$;

$a = -g = -9.80 \text{ m/s}^2$; and $t = 1.00 \text{ s}$. We also know from the solution above that $y_1 = 8.10 \text{ m}$.

2. Identify the best equation to use. The most straightforward is $v = v_0 - gt$ (from $v = v_0 + at$, where $a = \text{gravitational acceleration} = -g$).
3. Plug in the knowns and solve.

Equation:

$$v_1 = v_0 - gt = 13.0 \text{ m/s} - (9.80 \text{ m/s}^2)(1.00 \text{ s}) = 3.20 \text{ m/s}$$

Discussion

The positive value for v_1 means that the rock is still heading upward at $t = 1.00 \text{ s}$. However, it has slowed from its original 13.0 m/s , as expected.

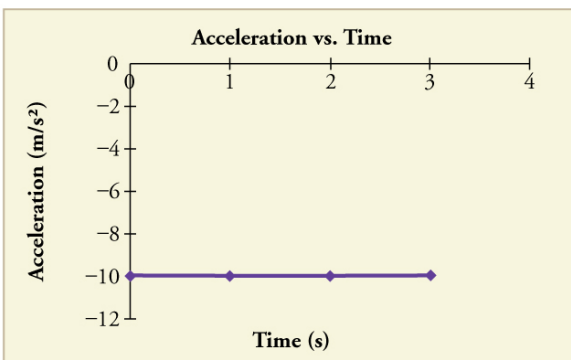
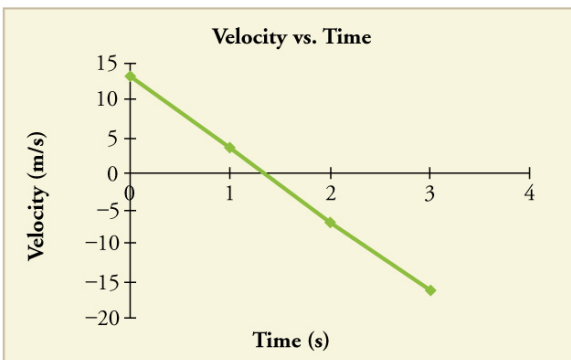
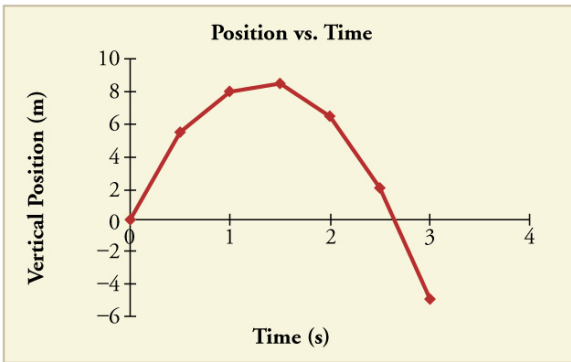
Solution for Remaining Times

The procedures for calculating the position and velocity at $t = 2.00 \text{ s}$ and 3.00 s are the same as those above. The results are summarized in [\[link\]](#) and illustrated in [\[link\]](#).

Time, t	Position, y	Velocity, v	Acceleration, a
1.00 s	8.10 m	3.20 m/s	-9.80 m/s^2
2.00 s	6.40 m	-6.60 m/s	-9.80 m/s^2
3.00 s	-5.10 m	-16.4 m/s	-9.80 m/s^2

Results

Graphing the data helps us understand it more clearly.



Vertical position, vertical velocity, and vertical acceleration vs. time for a rock thrown vertically up at the edge of a cliff. Notice that velocity changes linearly with time and that acceleration is constant.

Misconception Alert! Notice that the position vs. time graph shows vertical position only. It is easy to get the impression that the graph shows some

horizontal motion—the shape of the graph looks like the path of a projectile. But this is not the case; the horizontal axis is *time*, not space. The actual path of the rock in space is straight up, and straight down.

Discussion

The interpretation of these results is important. At 1.00 s the rock is above its starting point and heading upward, since y_1 and v_1 are both positive. At 2.00 s, the rock is still above its starting point, but the negative velocity means it is moving downward. At 3.00 s, both y_3 and v_3 are negative, meaning the rock is below its starting point and continuing to move downward. Notice that when the rock is at its highest point (at 1.5 s), its velocity is zero, but its acceleration is still -9.80 m/s^2 . Its acceleration is -9.80 m/s^2 for the whole trip—while it is moving up and while it is moving down. Note that the values for y are the positions (or displacements) of the rock, not the total distances traveled. Finally, note that free-fall applies to upward motion as well as downward. Both have the same acceleration—the acceleration due to gravity, which remains constant the entire time. Astronauts training in the famous Vomit Comet, for example, experience free-fall while arcing up as well as down, as we will discuss in more detail later.

Note:

Making Connections: Take-Home Experiment—Reaction Time

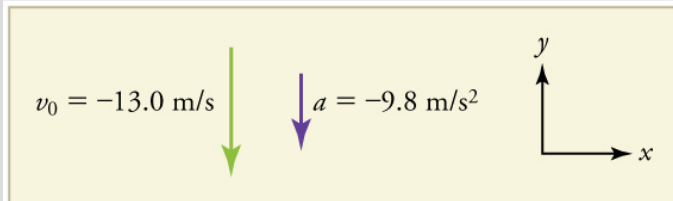
A simple experiment can be done to determine your reaction time. Have a friend hold a ruler between your thumb and index finger, separated by about 1 cm. Note the mark on the ruler that is right between your fingers. Have your friend drop the ruler unexpectedly, and try to catch it between your two fingers. Note the new reading on the ruler. Assuming acceleration is that due to gravity, calculate your reaction time. How far would you travel in a car (moving at 30 m/s) if the time it took your foot to go from the gas pedal to the brake was twice this reaction time?

Example:**Calculating Velocity of a Falling Object: A Rock Thrown Down**

What happens if the person on the cliff throws the rock straight down, instead of straight up? To explore this question, calculate the velocity of the rock when it is 5.10 m below the starting point, and has been thrown downward with an initial speed of 13.0 m/s.

Strategy

Draw a sketch.



Since up is positive, the final position of the rock will be negative because it finishes below the starting point at $y_0 = 0$. Similarly, the initial velocity is downward and therefore negative, as is the acceleration due to gravity. We expect the final velocity to be negative since the rock will continue to move downward.

Solution

1. Identify the knowns. $y_0 = 0$; $y_1 = -5.10$ m; $v_0 = -13.0$ m/s; $a = -g = -9.80$ m/s².
2. Choose the kinematic equation that makes it easiest to solve the problem. The equation $v^2 = v_0^2 + 2a(y - y_0)$ works well because the only unknown in it is v . (We will plug y_1 in for y .)
3. Enter the known values

Equation:

$$v^2 = (-13.0 \text{ m/s})^2 + 2(-9.80 \text{ m/s}^2)(-5.10 \text{ m} - 0 \text{ m}) = 268.96 \text{ m}^2/\text{s}^2,$$

where we have retained extra significant figures because this is an intermediate result.

Taking the square root, and noting that a square root can be positive or negative, gives

Equation:

$$v = \pm 16.4 \text{ m/s}.$$

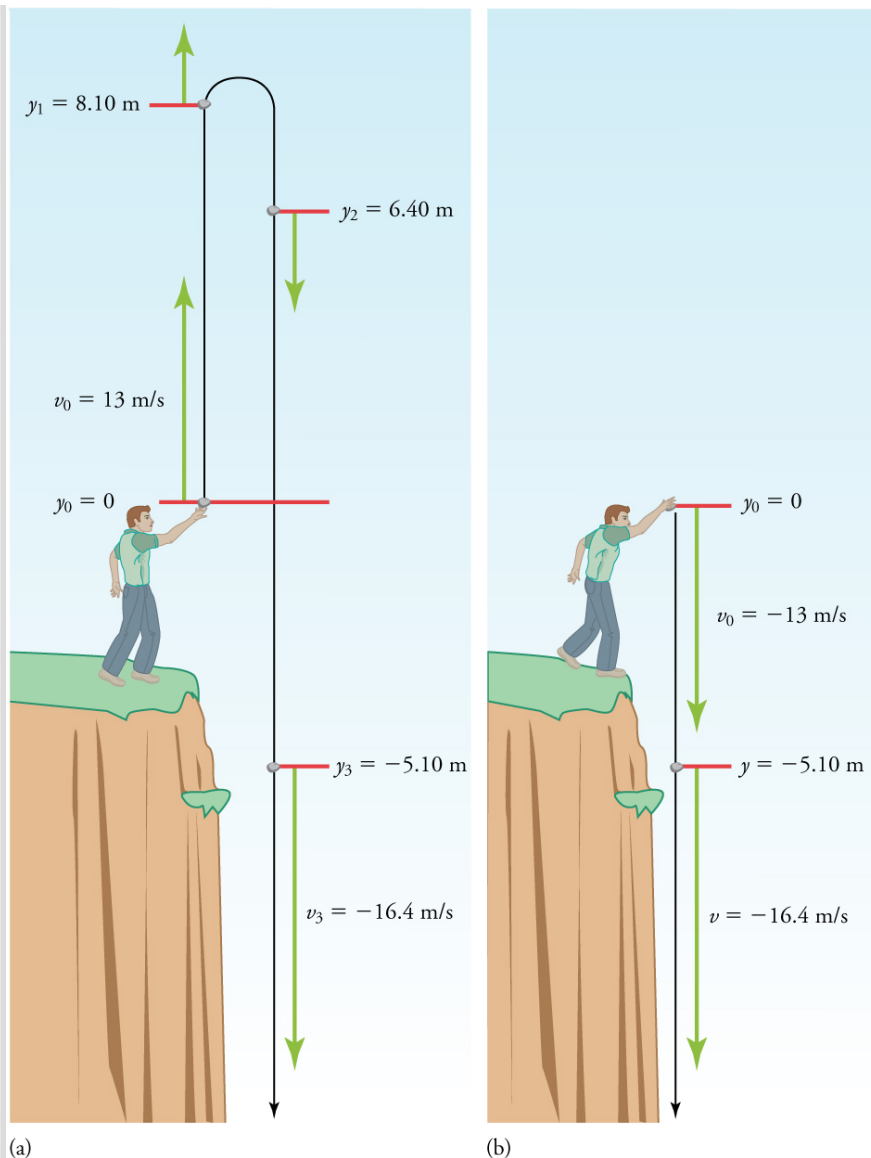
The negative root is chosen to indicate that the rock is still heading down. Thus,

Equation:

$$v = -16.4 \text{ m/s.}$$

Discussion

Note that *this is exactly the same velocity the rock had at this position when it was thrown straight upward with the same initial speed.* (See [\[link\]](#) and [\[link\]](#)(a).) This is not a coincidental result. Because we only consider the acceleration due to gravity in this problem, the *speed* of a falling object depends only on its initial speed and its vertical position relative to the starting point. For example, if the velocity of the rock is calculated at a height of 8.10 m above the starting point (using the method from [\[link\]](#)) when the initial velocity is 13.0 m/s straight up, a result of $\pm 3.20 \text{ m/s}$ is obtained. Here both signs are meaningful; the positive value occurs when the rock is at 8.10 m and heading up, and the negative value occurs when the rock is at 8.10 m and heading back down. It has the same *speed* but the opposite direction.



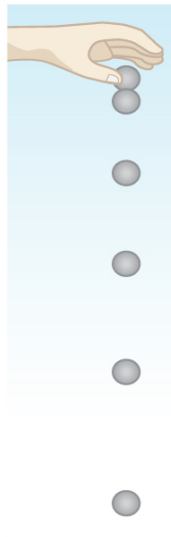
(a) A person throws a rock straight up, as explored in [\[link\]](#). The arrows are velocity vectors at 0, 1.00, 2.00, and 3.00 s. (b) A person throws a rock straight down from a cliff with the same initial speed as before, as in [\[link\]](#). Note that at the same distance below the point of release, the rock has the same velocity in both cases.

Another way to look at it is this: In [\[link\]](#), the rock is thrown up with an initial velocity of 13.0 m/s . It rises and then falls back down. When its

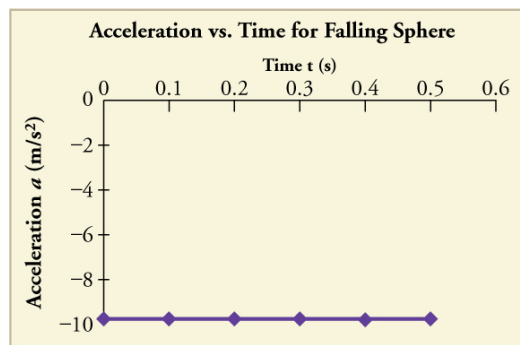
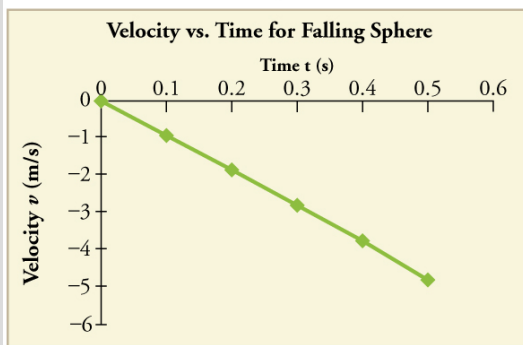
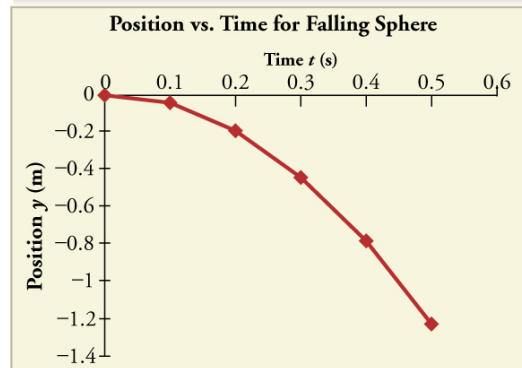
position is $y = 0$ on its way back down, its velocity is -13.0 m/s . That is, it has the same speed on its way down as on its way up. We would then expect its velocity at a position of $y = -5.10 \text{ m}$ to be the same whether we have thrown it upwards at $+13.0 \text{ m/s}$ or thrown it downwards at -13.0 m/s . The velocity of the rock on its way down from $y = 0$ is the same whether we have thrown it up or down to start with, as long as the speed with which it was initially thrown is the same.

Example:**Find g from Data on a Falling Object**

The acceleration due to gravity on Earth differs slightly from place to place, depending on topography (e.g., whether you are on a hill or in a valley) and subsurface geology (whether there is dense rock like iron ore as opposed to light rock like salt beneath you.) The precise acceleration due to gravity can be calculated from data taken in an introductory physics laboratory course. An object, usually a metal ball for which air resistance is negligible, is dropped and the time it takes to fall a known distance is measured. See, for example, [\[link\]](#). Very precise results can be produced with this method if sufficient care is taken in measuring the distance fallen and the elapsed time.



y (m)	v (m/s)	t (s)
0	0	0
-0.049	-0.98	0.1
-0.196	-1.96	0.2
-0.441	-2.94	0.3
-0.784	-3.92	0.4
-1.225	-4.90	0.5



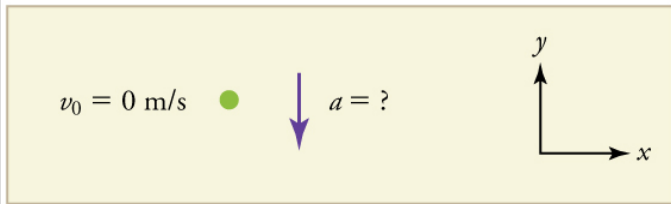
Positions and velocities of a metal ball released from rest when air resistance is negligible. Velocity is seen to increase linearly with time while displacement increases with time squared.

Acceleration is a constant and is equal to gravitational acceleration.

Suppose the ball falls 1.0000 m in 0.45173 s. Assuming the ball is not affected by air resistance, what is the precise acceleration due to gravity at this location?

Strategy

Draw a sketch.



We need to solve for acceleration a . Note that in this case, displacement is downward and therefore negative, as is acceleration.

Solution

1. Identify the knowns. $y_0 = 0$; $y = -1.0000$ m; $t = 0.45173$; $v_0 = 0$.
2. Choose the equation that allows you to solve for a using the known values.

Equation:

$$y = y_0 + v_0 t + \frac{1}{2} a t^2$$

3. Substitute 0 for v_0 and rearrange the equation to solve for a . Substituting 0 for v_0 yields

Equation:

$$y = y_0 + \frac{1}{2} a t^2.$$

Solving for a gives

Equation:

$$a = \frac{2(y - y_0)}{t^2}.$$

4. Substitute known values yields

Equation:

$$a = \frac{2(-1.0000 \text{ m} - 0)}{(0.45173 \text{ s})^2} = -9.8010 \text{ m/s}^2,$$

so, because $a = -g$ with the directions we have chosen,

Equation:

$$g = 9.8010 \text{ m/s}^2.$$

Discussion

The negative value for a indicates that the gravitational acceleration is downward, as expected. We expect the value to be somewhere around the average value of 9.80 m/s^2 , so 9.8010 m/s^2 makes sense. Since the data going into the calculation are relatively precise, this value for g is more precise than the average value of 9.80 m/s^2 ; it represents the local value for the acceleration due to gravity.

Exercise:

Check Your Understanding

Problem:

A chunk of ice breaks off a glacier and falls 30.0 meters before it hits the water. Assuming it falls freely (there is no air resistance), how long does it take to hit the water?

Solution:

We know that initial position $y_0 = 0$, final position $y = -30.0 \text{ m}$, and $a = -g = -9.80 \text{ m/s}^2$. We can then use the equation

$y = y_0 + v_0t + \frac{1}{2}at^2$ to solve for t . Inserting $a = -g$, we obtain

Equation:

$$y = 0 + 0 - \frac{1}{2}gt^2$$

$$t^2 = \frac{2y}{-g}$$

$$t = \pm \sqrt{\frac{2y}{-g}} = \pm \sqrt{\frac{2(-30.0 \text{ m})}{-9.80 \text{ m/s}^2}} = \pm \sqrt{6.12 \text{ s}^2} = 2.47 \text{ s} \approx 2.5 \text{ s}$$

where we take the positive value as the physically relevant answer. Thus, it takes about 2.5 seconds for the piece of ice to hit the water.

Note:

PhET Explorations: Equation Grapher

Learn about graphing polynomials. The shape of the curve changes as the constants are adjusted. View the curves for the individual terms (e.g. $y = bx$) to see how they add to generate the polynomial curve.

https://phet.colorado.edu/sims/equation-grapher/equation-grapher_en.html

Section Summary

- An object in free-fall experiences constant acceleration if air resistance is negligible.
- On Earth, all free-falling objects have an acceleration due to gravity g , which averages

Equation:

$$g = 9.80 \text{ m/s}^2.$$

- Whether the acceleration a should be taken as $+g$ or $-g$ is determined by your choice of coordinate system. If you choose the upward direction as positive, $a = -g = -9.80 \text{ m/s}^2$ is negative. In the opposite case, $a = +g = 9.80 \text{ m/s}^2$ is positive. Since acceleration is constant, the kinematic equations above can be applied with the appropriate $+g$ or $-g$ substituted for a .
- For objects in free-fall, up is normally taken as positive for displacement, velocity, and acceleration.

Conceptual Questions

Exercise:

Problem:

What is the acceleration of a rock thrown straight upward on the way up? At the top of its flight? On the way down?

Exercise:**Problem:**

An object that is thrown straight up falls back to Earth. This is one-dimensional motion. (a) When is its velocity zero? (b) Does its velocity change direction? (c) Does the acceleration due to gravity have the same sign on the way up as on the way down?

Exercise:**Problem:**

Suppose you throw a rock nearly straight up at a coconut in a palm tree, and the rock misses on the way up but hits the coconut on the way down. Neglecting air resistance, how does the speed of the rock when it hits the coconut on the way down compare with what it would have been if it had hit the coconut on the way up? Is it more likely to dislodge the coconut on the way up or down? Explain.

Exercise:**Problem:**

If an object is thrown straight up and air resistance is negligible, then its speed when it returns to the starting point is the same as when it was released. If air resistance were not negligible, how would its speed upon return compare with its initial speed? How would the maximum height to which it rises be affected?

Exercise:**Problem:**

The severity of a fall depends on your speed when you strike the ground. All factors but the acceleration due to gravity being the same, how many times higher could a safe fall on the Moon be than on Earth (gravitational acceleration on the Moon is about $1/6$ that of the Earth)?

Exercise:**Problem:**

How many times higher could an astronaut jump on the Moon than on Earth if his takeoff speed is the same in both locations (gravitational acceleration on the Moon is about $1/6$ of g on Earth)?

Problems & Exercises

Assume air resistance is negligible unless otherwise stated.

Exercise:**Problem:**

Calculate the displacement and velocity at times of (a) 0.500, (b) 1.00, (c) 1.50, and (d) 2.00 s for a ball thrown straight up with an initial velocity of 15.0 m/s. Take the point of release to be $y_0 = 0$.

Solution:

(a) $y_1 = 6.28 \text{ m}$; $v_1 = 10.1 \text{ m/s}$

(b) $y_2 = 10.1 \text{ m}$; $v_2 = 5.20 \text{ m/s}$

(c) $y_3 = 11.5 \text{ m}$; $v_3 = 0.300 \text{ m/s}$

(d) $y_4 = 10.4 \text{ m}$; $v_4 = -4.60 \text{ m/s}$

Exercise:**Problem:**

Calculate the displacement and velocity at times of (a) 0.500, (b) 1.00, (c) 1.50, (d) 2.00, and (e) 2.50 s for a rock thrown straight down with an initial velocity of 14.0 m/s from the Verrazano Narrows Bridge in New York City. The roadway of this bridge is 70.0 m above the water.

Exercise:

Problem:

A basketball referee tosses the ball straight up for the starting tip-off. At what velocity must a basketball player leave the ground to rise 1.25 m above the floor in an attempt to get the ball?

Solution:

$$v_0 = 4.95 \text{ m/s}$$

Exercise:**Problem:**

A rescue helicopter is hovering over a person whose boat has sunk. One of the rescuers throws a life preserver straight down to the victim with an initial velocity of 1.40 m/s and observes that it takes 1.8 s to reach the water. (a) List the knowns in this problem. (b) How high above the water was the preserver released? Note that the downdraft of the helicopter reduces the effects of air resistance on the falling life preserver, so that an acceleration equal to that of gravity is reasonable.

Exercise:**Problem:**

A dolphin in an aquatic show jumps straight up out of the water at a velocity of 13.0 m/s. (a) List the knowns in this problem. (b) How high does his body rise above the water? To solve this part, first note that the final velocity is now a known and identify its value. Then identify the unknown, and discuss how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, checking units, and discuss whether the answer is reasonable. (c) How long is the dolphin in the air? Neglect any effects due to his size or orientation.

Solution:

$$(a) a = -9.80 \text{ m/s}^2; v_0 = 13.0 \text{ m/s}; y_0 = 0 \text{ m}$$

(b) $v = 0\text{ m/s}$. Unknown is distance y to top of trajectory, where velocity is zero. Use equation $v^2 = v_0^2 + 2a(y - y_0)$ because it contains all known values except for y , so we can solve for y . Solving for y gives

Equation:

$$v^2 - v_0^2 = 2a(y - y_0)$$

$$\frac{v^2 - v_0^2}{2a} = y - y_0$$

$$y = y_0 + \frac{v^2 - v_0^2}{2a} = 0\text{ m} + \frac{(0\text{ m/s})^2 - (13.0\text{ m/s})^2}{2(-9.80\text{ m/s}^2)} = 8.62\text{ m}$$

Dolphins measure about 2 meters long and can jump several times their length out of the water, so this is a reasonable result.

(c) 2.65 s

Exercise:

Problem:

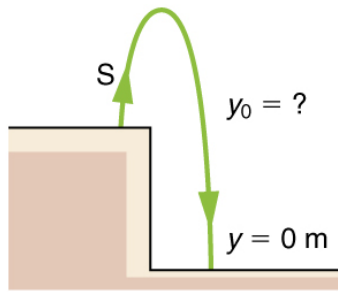
A swimmer bounces straight up from a diving board and falls feet first into a pool. She starts with a velocity of 4.00 m/s, and her takeoff point is 1.80 m above the pool. (a) How long are her feet in the air? (b) What is her highest point above the board? (c) What is her velocity when her feet hit the water?

Exercise:

Problem:

(a) Calculate the height of a cliff if it takes 2.35 s for a rock to hit the ground when it is thrown straight up from the cliff with an initial velocity of 8.00 m/s. (b) How long would it take to reach the ground if it is thrown straight down with the same speed?

Solution:



(a) 8.26 m

(b) 0.717 s

Exercise:

Problem:

A very strong, but inept, shot putter puts the shot straight up vertically with an initial velocity of 11.0 m/s. How long does he have to get out of the way if the shot was released at a height of 2.20 m, and he is 1.80 m tall?

Exercise:

Problem:

You throw a ball straight up with an initial velocity of 15.0 m/s. It passes a tree branch on the way up at a height of 7.00 m. How much additional time will pass before the ball passes the tree branch on the way back down?

Solution:

1.91 s

Exercise:

Problem:

A kangaroo can jump over an object 2.50 m high. (a) Calculate its vertical speed when it leaves the ground. (b) How long is it in the air?

Exercise:

Problem:

Standing at the base of one of the cliffs of Mt. Arapiles in Victoria, Australia, a hiker hears a rock break loose from a height of 105 m. He can't see the rock right away but then does, 1.50 s later. (a) How far above the hiker is the rock when he can see it? (b) How much time does he have to move before the rock hits his head?

Solution:

(a) 94.0 m

(b) 3.13 s

Exercise:**Problem:**

An object is dropped from a height of 75.0 m above ground level. (a) Determine the distance traveled during the first second. (b) Determine the final velocity at which the object hits the ground. (c) Determine the distance traveled during the last second of motion before hitting the ground.

Exercise:**Problem:**

There is a 250-m-high cliff at Half Dome in Yosemite National Park in California. Suppose a boulder breaks loose from the top of this cliff. (a) How fast will it be going when it strikes the ground? (b) Assuming a reaction time of 0.300 s, how long will a tourist at the bottom have to get out of the way after hearing the sound of the rock breaking loose (neglecting the height of the tourist, which would become negligible anyway if hit)? The speed of sound is 335 m/s on this day.

Solution:

(a) -70.0 m/s (downward)

(b) 6.10 s

Exercise:**Problem:**

A ball is thrown straight up. It passes a 2.00-m-high window 7.50 m off the ground on its path up and takes 0.312 s to go past the window. What was the ball's initial velocity? Hint: First consider only the distance along the window, and solve for the ball's velocity at the bottom of the window. Next, consider only the distance from the ground to the bottom of the window, and solve for the initial velocity using the velocity at the bottom of the window as the final velocity.

Exercise:**Problem:**

Suppose you drop a rock into a dark well and, using precision equipment, you measure the time for the sound of a splash to return. (a) Neglecting the time required for sound to travel up the well, calculate the distance to the water if the sound returns in 2.0000 s. (b) Now calculate the distance taking into account the time for sound to travel up the well. The speed of sound is 332.00 m/s in this well.

Solution:

(a) 19.6 m

(b) 18.5 m

Exercise:**Problem:**

A steel ball is dropped onto a hard floor from a height of 1.50 m and rebounds to a height of 1.45 m. (a) Calculate its velocity just before it strikes the floor. (b) Calculate its velocity just after it leaves the floor on its way back up. (c) Calculate its acceleration during contact with the floor if that contact lasts 0.0800 ms (8.00×10^{-5} s). (d) How much did the ball compress during its collision with the floor, assuming the floor is absolutely rigid?

Exercise:

Problem:

A coin is dropped from a hot-air balloon that is 300 m above the ground and rising at 10.0 m/s upward. For the coin, find (a) the maximum height reached, (b) its position and velocity 4.00 s after being released, and (c) the time before it hits the ground.

Solution:

(a) 305 m

(b) 262 m, -29.2 m/s

(c) 8.91 s

Exercise:**Problem:**

A soft tennis ball is dropped onto a hard floor from a height of 1.50 m and rebounds to a height of 1.10 m. (a) Calculate its velocity just before it strikes the floor. (b) Calculate its velocity just after it leaves the floor on its way back up. (c) Calculate its acceleration during contact with the floor if that contact lasts 3.50 ms (3.50×10^{-3} s). (d) How much did the ball compress during its collision with the floor, assuming the floor is absolutely rigid?

Glossary

free-fall

the state of movement that results from gravitational force only

acceleration due to gravity

acceleration of an object as a result of gravity

Graphical Analysis of One-Dimensional Motion

- Describe a straight-line graph in terms of its slope and y-intercept.
- Determine average velocity or instantaneous velocity from a graph of position vs. time.
- Determine average or instantaneous acceleration from a graph of velocity vs. time.
- Derive a graph of velocity vs. time from a graph of position vs. time.
- Derive a graph of acceleration vs. time from a graph of velocity vs. time.

A graph, like a picture, is worth a thousand words. Graphs not only contain numerical information; they also reveal relationships between physical quantities. This section uses graphs of position, velocity, and acceleration versus time to illustrate one-dimensional kinematics.

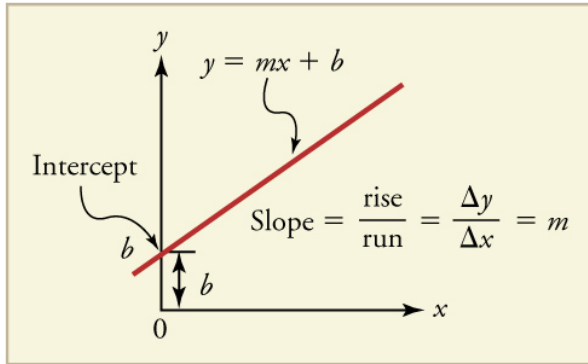
Slopes and General Relationships

First note that graphs in this text have perpendicular axes, one horizontal and the other vertical. When two physical quantities are plotted against one another in such a graph, the horizontal axis is usually considered to be an **independent variable** and the vertical axis a **dependent variable**. If we call the horizontal axis the x -axis and the vertical axis the y -axis, as in [\[link\]](#), a straight-line graph has the general form

Equation:

$$y = mx + b.$$

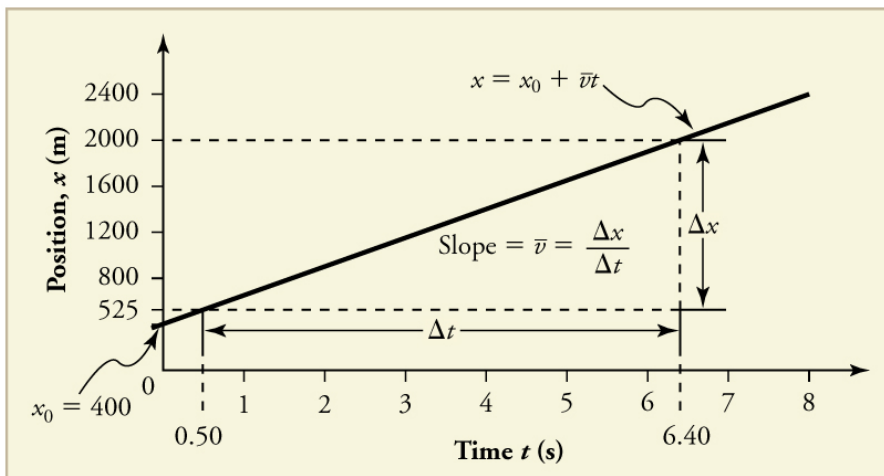
Here m is the **slope**, defined to be the rise divided by the run (as seen in the figure) of the straight line. The letter b is used for the **y-intercept**, which is the point at which the line crosses the vertical axis.



A straight-line graph. The equation for a straight line is $y = mx + b$.

Graph of Position vs. Time ($a = 0$, so v is constant)

Time is usually an independent variable that other quantities, such as position, depend upon. A graph of position versus time would, thus, have x on the vertical axis and t on the horizontal axis. [\[link\]](#) is just such a straight-line graph. It shows a graph of position versus time for a jet-powered car on a very flat dry lake bed in Nevada.



Graph of position versus time for a jet-powered car on the Bonneville Salt Flats.

Using the relationship between dependent and independent variables, we see that the slope in the graph above is average velocity \bar{v} and the intercept is position at time zero—that is, x_0 . Substituting these symbols into $y = mx + b$ gives

Equation:

$$x = \bar{v}t + x_0$$

or

Equation:

$$x = x_0 + \bar{v}t.$$

Thus a graph of position versus time gives a general relationship among displacement(change in position), velocity, and time, as well as giving detailed numerical information about a specific situation.

Note:

The Slope of x vs. t

The slope of the graph of position x vs. time t is velocity v .

Equation:

$$\text{slope} = \frac{\Delta x}{\Delta t} = v$$

Notice that this equation is the same as that derived algebraically from other motion equations in [Motion Equations for Constant Acceleration in One Dimension](#).

From the figure we can see that the car has a position of 25 m at 0.50 s and 2000 m at 6.40 s. Its position at other times can be read from the graph; furthermore, information about its velocity and acceleration can also be obtained from the graph.

Example:

**Determining Average Velocity from a Graph of Position versus Time:
Jet Car**

Find the average velocity of the car whose position is graphed in [\[link\]](#).

Strategy

The slope of a graph of x vs. t is average velocity, since slope equals rise over run. In this case, rise = change in position and run = change in time, so that

Equation:

$$\text{slope} = \frac{\Delta x}{\Delta t} = \bar{v}.$$

Since the slope is constant here, any two points on the graph can be used to find the slope. (Generally speaking, it is most accurate to use two widely separated points on the straight line. This is because any error in reading data from the graph is proportionally smaller if the interval is larger.)

Solution

1. Choose two points on the line. In this case, we choose the points labeled on the graph: (6.4 s, 2000 m) and (0.50 s, 525 m). (Note, however, that you could choose any two points.)

2. Substitute the x and t values of the chosen points into the equation. Remember in calculating change (Δ) we always use final value minus initial value.

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{2000 \text{ m} - 525 \text{ m}}{6.4 \text{ s} - 0.50 \text{ s}},$$

yielding

Equation:

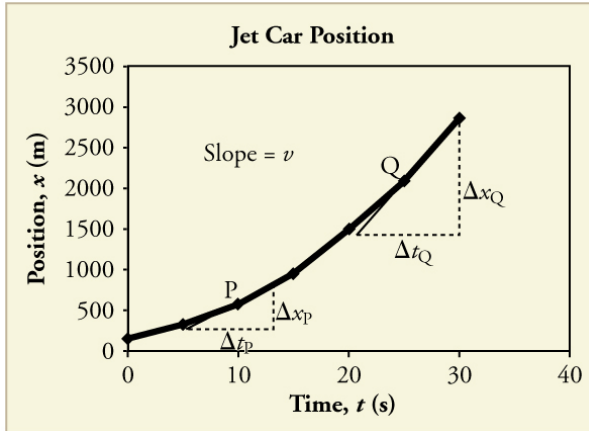
$$\bar{v} = 250 \text{ m/s.}$$

Discussion

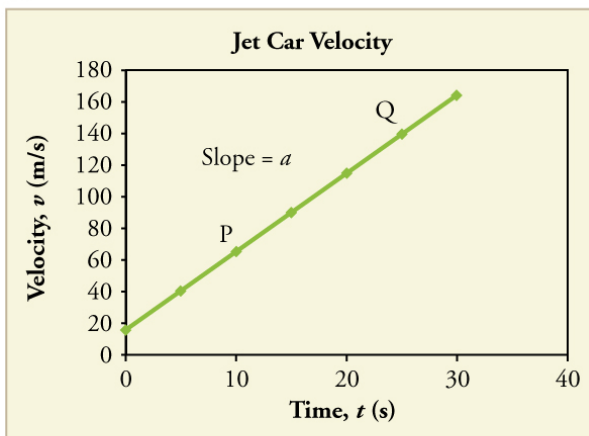
This is an impressively large land speed (900 km/h, or about 560 mi/h): much greater than the typical highway speed limit of 60 mi/h (27 m/s or 96 km/h), but considerably shy of the record of 343 m/s (1234 km/h or 766 mi/h) set in 1997.

Graphs of Motion when a is constant but $a \neq 0$

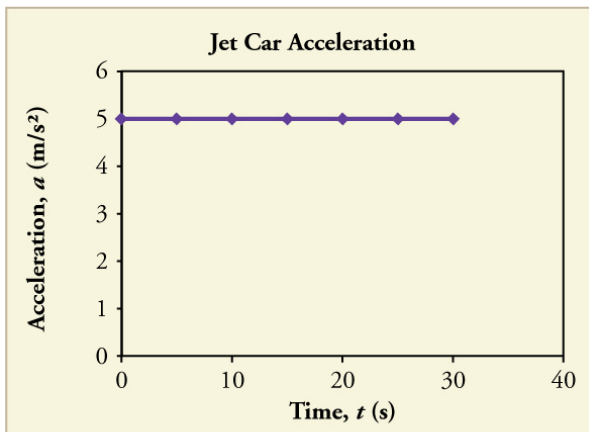
The graphs in [\[link\]](#) below represent the motion of the jet-powered car as it accelerates toward its top speed, but only during the time when its acceleration is constant. Time starts at zero for this motion (as if measured with a stopwatch), and the position and velocity are initially 200 m and 15 m/s, respectively.



(a)



(b)



(c)

Graphs of motion of a jet-powered car during the time span when its acceleration is constant. (a) The slope of an x vs. t graph is velocity. This is

shown at two points, and the instantaneous velocities obtained are plotted in the next graph. Instantaneous velocity at any point is the slope of the tangent at that point. (b) The slope of the v vs. t graph is constant for this part of the motion, indicating constant acceleration. (c) Acceleration has the constant value of 5.0 m/s^2 over the time interval plotted.



A U.S. Air Force jet car speeds down a track. (credit: Matt Trostle, Flickr)

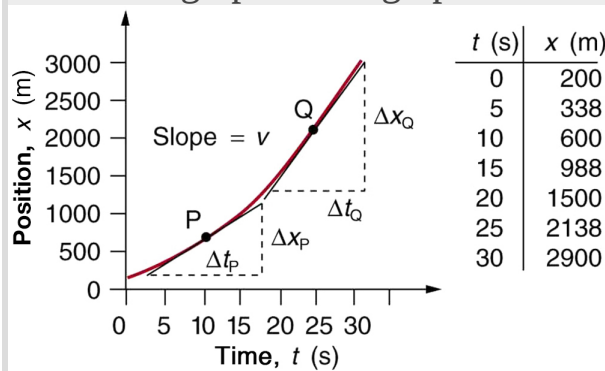
The graph of position versus time in [\[link\]](#)(a) is a curve rather than a straight line. The slope of the curve becomes steeper as time progresses,

showing that the velocity is increasing over time. The slope at any point on a position-versus-time graph is the instantaneous velocity at that point. It is found by drawing a straight line tangent to the curve at the point of interest and taking the slope of this straight line. Tangent lines are shown for two points in [\[link\]\(a\)](#). If this is done at every point on the curve and the values are plotted against time, then the graph of velocity versus time shown in [\[link\]\(b\)](#) is obtained. Furthermore, the slope of the graph of velocity versus time is acceleration, which is shown in [\[link\]\(c\)](#).

Example:

Determining Instantaneous Velocity from the Slope at a Point: Jet Car

Calculate the velocity of the jet car at a time of 25 s by finding the slope of the x vs. t graph in the graph below.



The slope of an x vs. t graph is velocity. This is shown at two points. Instantaneous velocity at any point is the slope of the tangent at that point.

Strategy

The slope of a curve at a point is equal to the slope of a straight line tangent to the curve at that point. This principle is illustrated in [\[link\]](#), where Q is the point at $t = 25$ s.

Solution

1. Find the tangent line to the curve at $t = 25$ s.

2. Determine the endpoints of the tangent. These correspond to a position of 1300 m at time 19 s and a position of 3120 m at time 32 s.
3. Plug these endpoints into the equation to solve for the slope, v .

Equation:

$$\text{slope} = v_Q = \frac{\Delta x_Q}{\Delta t_Q} = \frac{(3120 \text{ m} - 1300 \text{ m})}{(32 \text{ s} - 19 \text{ s})}$$

Thus,

Equation:

$$v_Q = \frac{1820 \text{ m}}{13 \text{ s}} = 140 \text{ m/s}.$$

Discussion

This is the value given in this figure's table for v at $t = 25$ s. The value of 140 m/s for v_Q is plotted in [\[link\]](#). The entire graph of v vs. t can be obtained in this fashion.

Carrying this one step further, we note that the slope of a velocity versus time graph is acceleration. Slope is rise divided by run; on a v vs. t graph, rise = change in velocity Δv and run = change in time Δt .

Note:

The Slope of v vs. t

The slope of a graph of velocity v vs. time t is acceleration a .

Equation:

$$\text{slope} = \frac{\Delta v}{\Delta t} = a$$

Since the velocity versus time graph in [\[link\]](#)(b) is a straight line, its slope is the same everywhere, implying that acceleration is constant. Acceleration versus time is graphed in [\[link\]](#)(c).

Additional general information can be obtained from [\[link\]](#) and the expression for a straight line, $y = mx + b$.

In this case, the vertical axis y is V , the intercept b is v_0 , the slope m is a , and the horizontal axis x is t . Substituting these symbols yields

Equation:

$$v = v_0 + at.$$

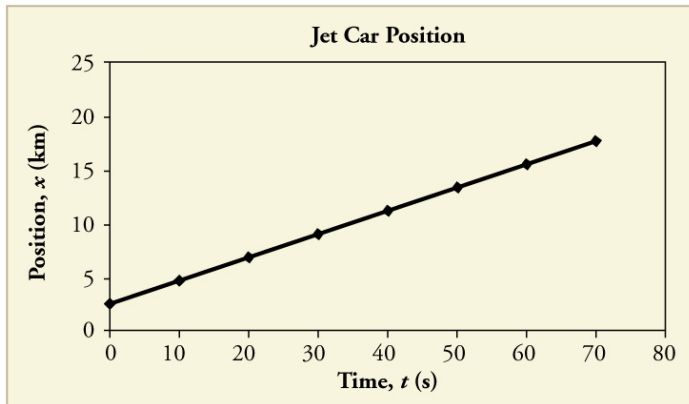
A general relationship for velocity, acceleration, and time has again been obtained from a graph. Notice that this equation was also derived algebraically from other motion equations in [Motion Equations for Constant Acceleration in One Dimension](#).

It is not accidental that the same equations are obtained by graphical analysis as by algebraic techniques. In fact, an important way to *discover* physical relationships is to measure various physical quantities and then make graphs of one quantity against another to see if they are correlated in any way. Correlations imply physical relationships and might be shown by smooth graphs such as those above. From such graphs, mathematical relationships can sometimes be postulated. Further experiments are then performed to determine the validity of the hypothesized relationships.

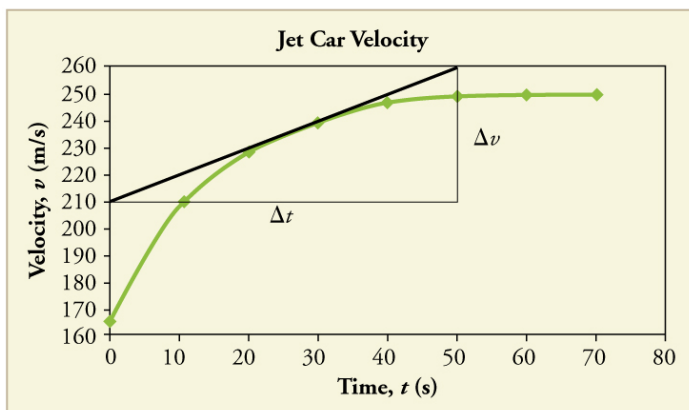
Graphs of Motion Where Acceleration is Not Constant

Now consider the motion of the jet car as it goes from 165 m/s to its top velocity of 250 m/s, graphed in [\[link\]](#). Time again starts at zero, and the initial position and velocity are 2900 m and 165 m/s, respectively. (These were the final position and velocity of the car in the motion graphed in [\[link\]](#).) Acceleration gradually decreases from 5.0 m/s^2 to zero when the car hits 250 m/s. The slope of the x vs. t graph increases until $t = 55 \text{ s}$, after which time the slope is constant. Similarly, velocity increases until 55

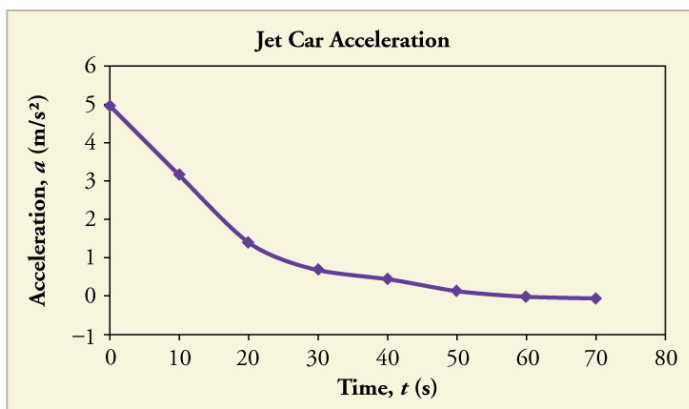
s and then becomes constant, since acceleration decreases to zero at 55 s and remains zero afterward.



(a)



(b)



(c)

Graphs of motion of a jet-powered car as it reaches its top velocity. This motion begins where the motion in

[\[link\]](#) ends. (a) The slope of this graph is velocity; it is plotted in the next graph. (b) The velocity gradually approaches its top value. The slope of this graph is acceleration; it is plotted in the final graph. (c) Acceleration gradually declines to zero when velocity becomes constant.

Example:**Calculating Acceleration from a Graph of Velocity versus Time**

Calculate the acceleration of the jet car at a time of 25 s by finding the slope of the v vs. t graph in [\[link\]](#)(b).

Strategy

The slope of the curve at $t = 25$ s is equal to the slope of the line tangent at that point, as illustrated in [\[link\]](#)(b).

Solution

Determine endpoints of the tangent line from the figure, and then plug them into the equation to solve for slope, a .

Equation:

$$\text{slope} = \frac{\Delta v}{\Delta t} = \frac{(260 \text{ m/s} - 210 \text{ m/s})}{(51 \text{ s} - 1.0 \text{ s})}$$

Equation:

$$a = \frac{50 \text{ m/s}}{50 \text{ s}} = 1.0 \text{ m/s}^2.$$

Discussion

Note that this value for a is consistent with the value plotted in [\[link\]](#)(c) at $t = 25$ s.

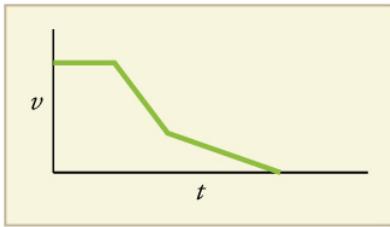
A graph of position versus time can be used to generate a graph of velocity versus time, and a graph of velocity versus time can be used to generate a graph of acceleration versus time. We do this by finding the slope of the graphs at every point. If the graph is linear (i.e., a line with a constant slope), it is easy to find the slope at any point and you have the slope for every point. Graphical analysis of motion can be used to describe both specific and general characteristics of kinematics. Graphs can also be used for other topics in physics. An important aspect of exploring physical relationships is to graph them and look for underlying relationships.

Exercise:

Check Your Understanding

Problem:

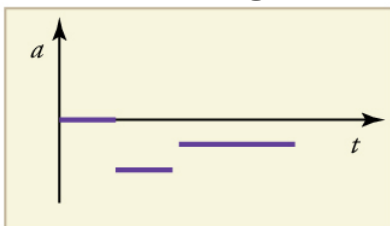
A graph of velocity vs. time of a ship coming into a harbor is shown below. (a) Describe the motion of the ship based on the graph. (b) What would a graph of the ship's acceleration look like?



Solution:

(a) The ship moves at constant velocity and then begins to decelerate at a constant rate. At some point, its deceleration rate decreases. It maintains this lower deceleration rate until it stops moving.

(b) A graph of acceleration vs. time would show zero acceleration in the first leg, large and constant negative acceleration in the second leg, and constant negative acceleration.



Section Summary

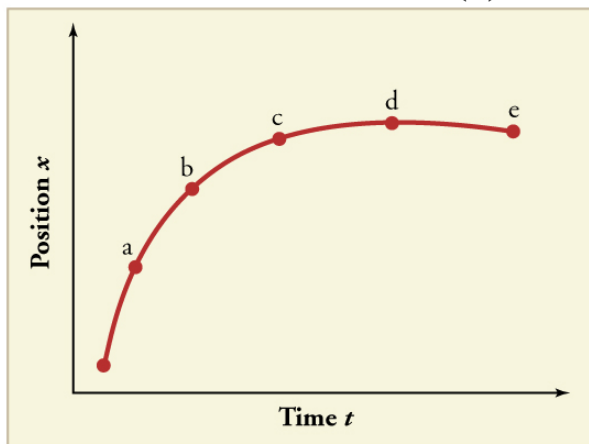
- Graphs of motion can be used to analyze motion.
- Graphical solutions yield identical solutions to mathematical methods for deriving motion equations.
- The slope of a graph of displacement x vs. time t is velocity v .
- The slope of a graph of velocity v vs. time t graph is acceleration a .
- Average velocity, instantaneous velocity, and acceleration can all be obtained by analyzing graphs.

Conceptual Questions

Exercise:

Problem:

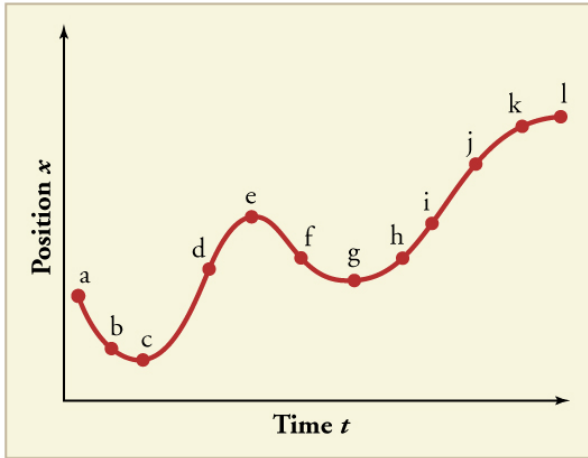
(a) Explain how you can use the graph of position versus time in [\[link\]](#) to describe the change in velocity over time. Identify (b) the time (t_a , t_b , t_c , t_d , or t_e) at which the instantaneous velocity is greatest, (c) the time at which it is zero, and (d) the time at which it is negative.



Exercise:

Problem:

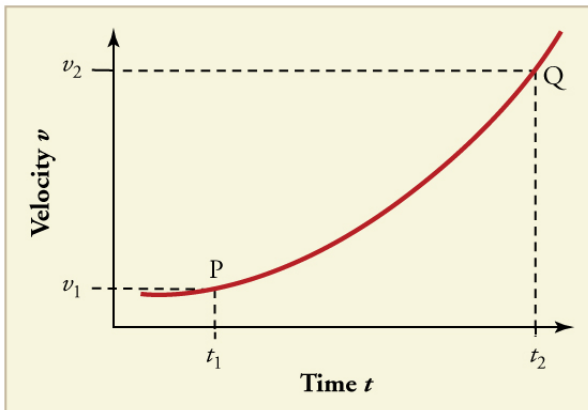
(a) Sketch a graph of velocity versus time corresponding to the graph of position versus time given in [\[link\]](#). (b) Identify the time or times (t_a , t_b , t_c , etc.) at which the instantaneous velocity is greatest. (c) At which times is it zero? (d) At which times is it negative?



Exercise:

Problem:

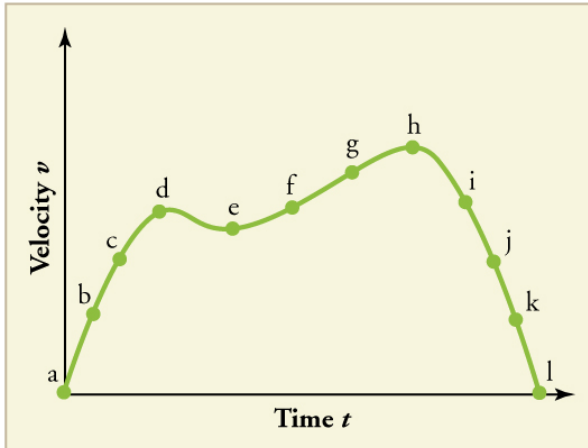
(a) Explain how you can determine the acceleration over time from a velocity versus time graph such as the one in [\[link\]](#). (b) Based on the graph, how does acceleration change over time?



Exercise:

Problem:

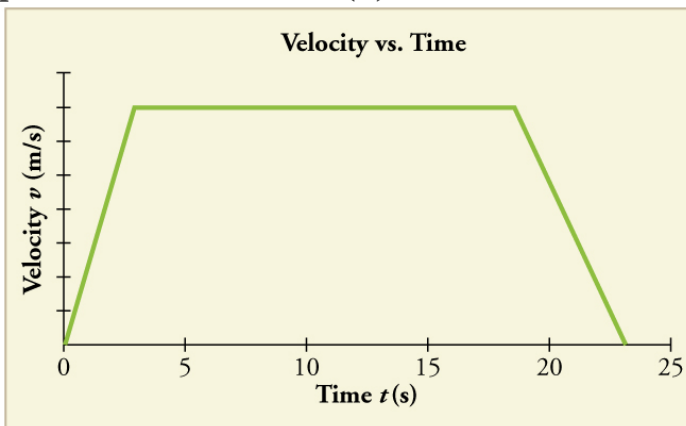
(a) Sketch a graph of acceleration versus time corresponding to the graph of velocity versus time given in [\[link\]](#). (b) Identify the time or times (t_a , t_b , t_c , etc.) at which the acceleration is greatest. (c) At which times is it zero? (d) At which times is it negative?



Exercise:

Problem:

Consider the velocity vs. time graph of a person in an elevator shown in [\[link\]](#). Suppose the elevator is initially at rest. It then accelerates for 3 seconds, maintains that velocity for 15 seconds, then decelerates for 5 seconds until it stops. The acceleration for the entire trip is not constant so we cannot use the equations of motion from [Motion Equations for Constant Acceleration in One Dimension](#) for the complete trip. (We could, however, use them in the three individual sections where acceleration is a constant.) Sketch graphs of (a) position vs. time and (b) acceleration vs. time for this trip.



Exercise:

Problem:

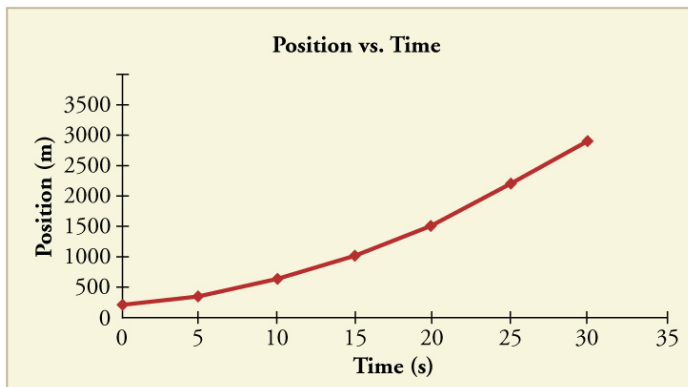
A cylinder is given a push and then rolls up an inclined plane. If the origin is the starting point, sketch the position, velocity, and acceleration of the cylinder vs. time as it goes up and then down the plane.

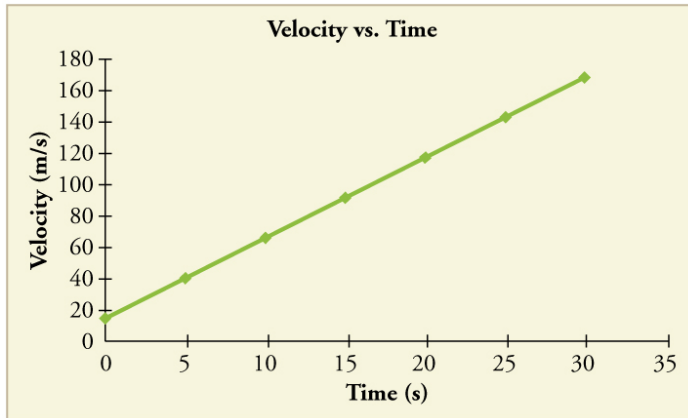
Problems & Exercises

Note: There is always uncertainty in numbers taken from graphs. If your answers differ from expected values, examine them to see if they are within data extraction uncertainties estimated by you.

Exercise:**Problem:**

(a) By taking the slope of the curve in [\[link\]](#), verify that the velocity of the jet car is 115 m/s at $t = 20$ s. (b) By taking the slope of the curve at any point in [\[link\]](#), verify that the jet car's acceleration is 5.0 m/s^2 .





Solution:

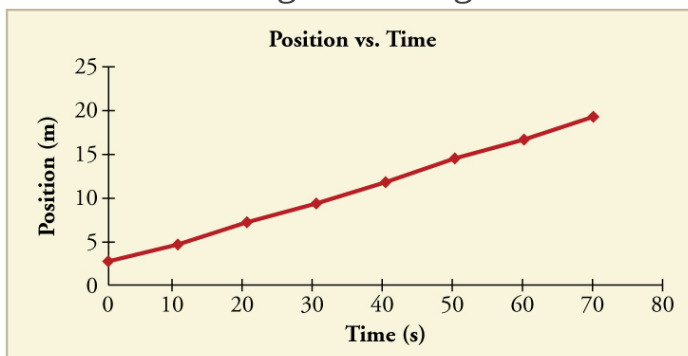
(a) 115 m/s

(b) 5.0 m/s²

Exercise:

Problem:

Using approximate values, calculate the slope of the curve in [\[link\]](#) to verify that the velocity at $t = 10.0$ s is 0.208 m/s. Assume all values are known to 3 significant figures.



Exercise:

Problem:

Using approximate values, calculate the slope of the curve in [\[link\]](#) to verify that the velocity at $t = 30.0$ s is approximately 0.24 m/s.

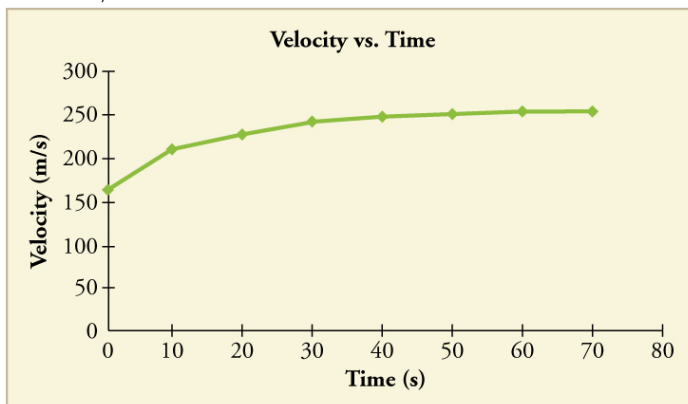
Solution:
Equation:

$$v = \frac{(11.7 - 6.95) \times 10^3 \text{ m}}{(40.0 - 20.0) \text{ s}} = 238 \text{ m/s}$$

Exercise:

Problem:

By taking the slope of the curve in [\[link\]](#), verify that the acceleration is 3.2 m/s^2 at $t = 10 \text{ s}$.

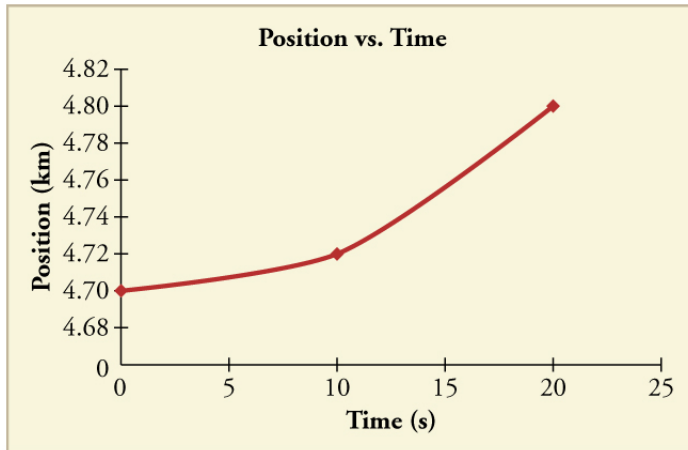


Exercise:

Problem:

Construct the position graph for the subway shuttle train as shown in [\[link\]](#)(a). Your graph should show the position of the train, in kilometers, from $t = 0$ to 20 s . You will need to use the information on acceleration and velocity given in the examples for this figure.

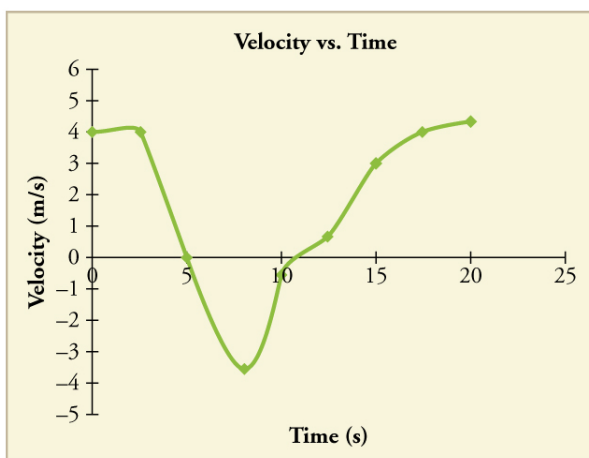
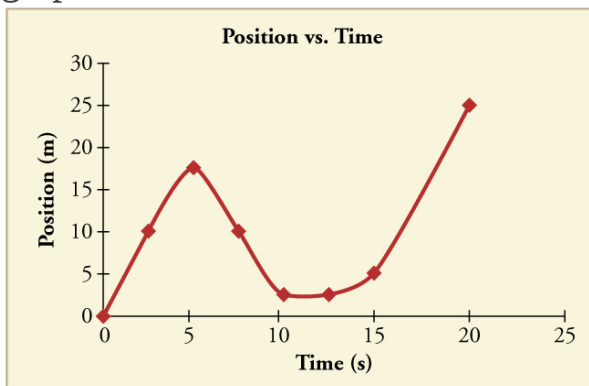
Solution:

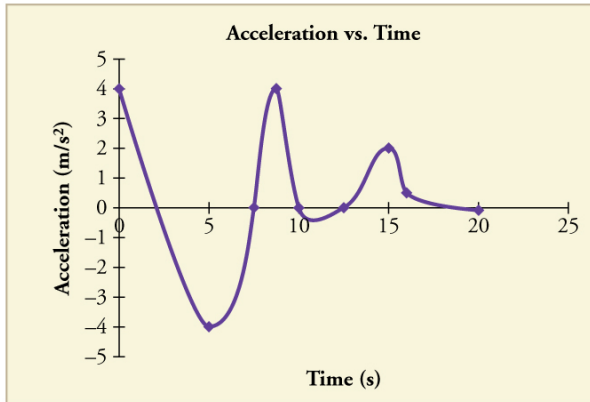


Exercise:

Problem:

(a) Take the slope of the curve in [\[link\]](#) to find the jogger's velocity at $t = 2.5$ s. (b) Repeat at 7.5 s. These values must be consistent with the graph in [\[link\]](#).

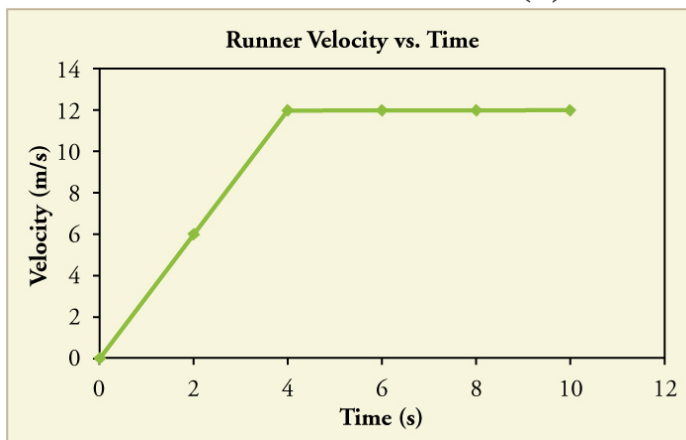




Exercise:

Problem:

A graph of $v(t)$ is shown for a world-class track sprinter in a 100-m race. (See [link](#)). (a) What is his average velocity for the first 4 s? (b) What is his instantaneous velocity at $t = 5$ s? (c) What is his average acceleration between 0 and 4 s? (d) What is his time for the race?



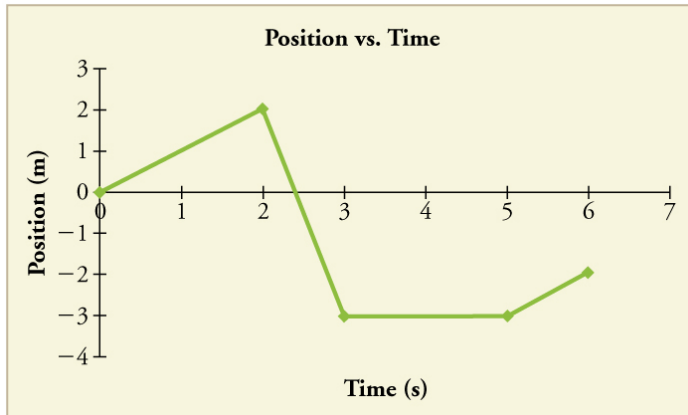
Solution:

- (a) 6 m/s
- (b) 12 m/s
- (c) 3 m/s^2
- (d) 10 s

Exercise:

Problem:

[\[link\]](#) shows the position graph for a particle for 5 s. Draw the corresponding velocity and acceleration graphs.



Glossary

independent variable

the variable that the dependent variable is measured with respect to;
usually plotted along the x -axis

dependent variable

the variable that is being measured; usually plotted along the y -axis

slope

the difference in y -value (the rise) divided by the difference in x -value
(the run) of two points on a straight line

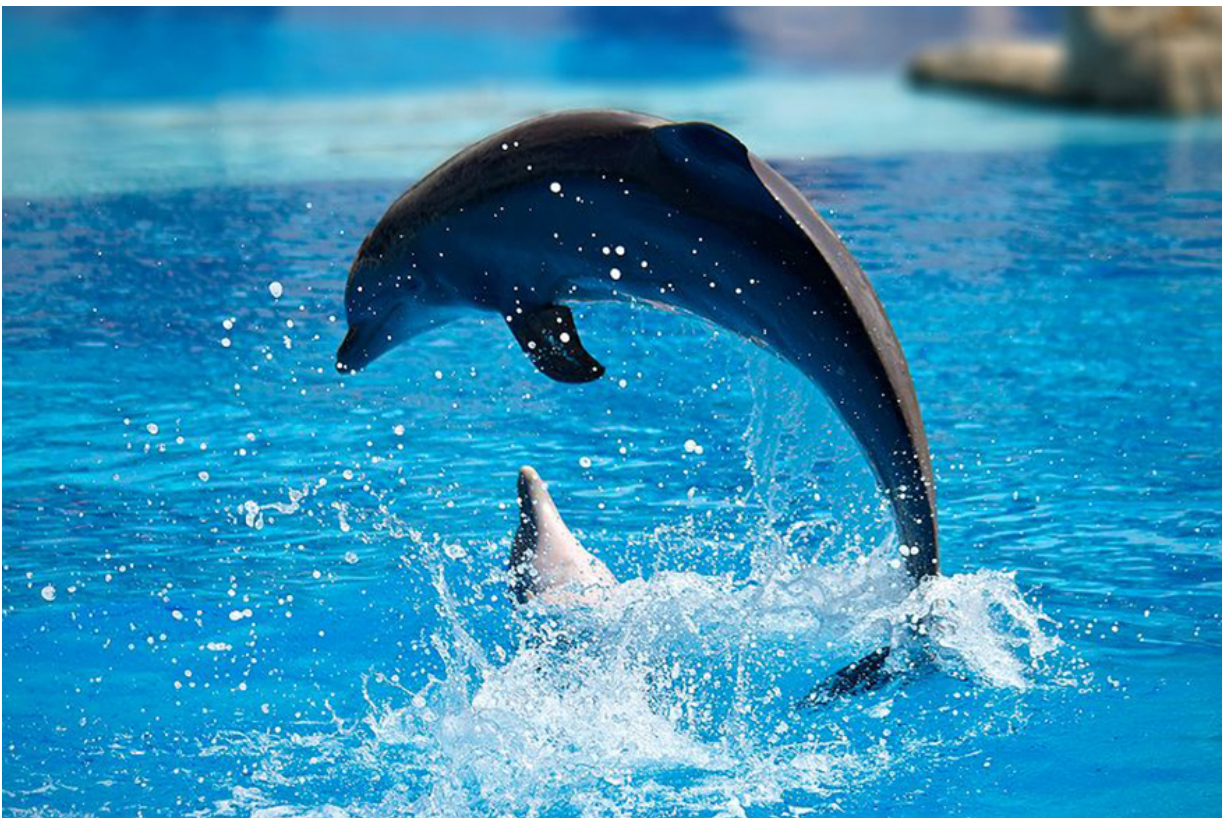
y -intercept

the y -value when $x = 0$, or when the graph crosses the y -axis

Introduction to Dynamics: Newton's Laws of Motion

class="introduction"

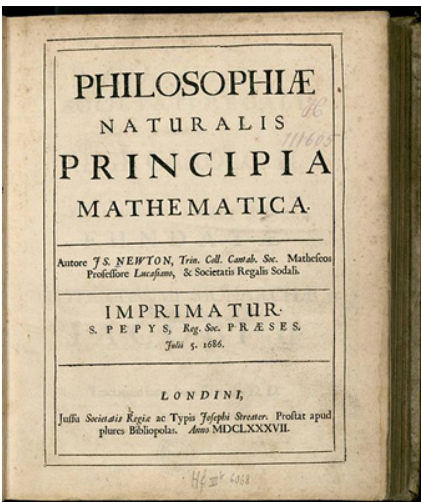
Newton's laws of motion describe the motion of the dolphin's path.
(credit: Jin Jang)



Motion draws our attention. Motion itself can be beautiful, causing us to marvel at the forces needed to achieve spectacular motion, such as that of a

dolphin jumping out of the water, or a pole vaulter, or the flight of a bird, or the orbit of a satellite. The study of motion is kinematics, but kinematics only *describes* the way objects move—their velocity and their acceleration. **Dynamics** considers the forces that affect the motion of moving objects and systems. Newton's laws of motion are the foundation of dynamics. These laws provide an example of the breadth and simplicity of principles under which nature functions. They are also universal laws in that they apply to similar situations on Earth as well as in space.

Isaac Newton's (1642–1727) laws of motion were just one part of the monumental work that has made him legendary. The development of Newton's laws marks the transition from the Renaissance into the modern era. This transition was characterized by a revolutionary change in the way people thought about the physical universe. For many centuries natural philosophers had debated the nature of the universe based largely on certain rules of logic with great weight given to the thoughts of earlier classical philosophers such as Aristotle (384–322 BC). Among the many great thinkers who contributed to this change were Newton and Galileo.



Isaac Newton's
monumental work,
*Philosophiæ
Naturalis Principia
Mathematica*, was
published in 1687. It
proposed scientific

laws that are still
used today to
describe the motion
of objects. (credit:
Service commun de
la documentation de
l'Université de
Strasbourg)

Galileo was instrumental in establishing *observation* as the absolute determinant of truth, rather than “logical” argument. Galileo’s use of the telescope was his most notable achievement in demonstrating the importance of observation. He discovered moons orbiting Jupiter and made other observations that were inconsistent with certain ancient ideas and religious dogma. For this reason, and because of the manner in which he dealt with those in authority, Galileo was tried by the Inquisition and punished. He spent the final years of his life under a form of house arrest. Because others before Galileo had also made discoveries by *observing* the nature of the universe, and because repeated observations verified those of Galileo, his work could not be suppressed or denied. After his death, his work was verified by others, and his ideas were eventually accepted by the church and scientific communities.

Galileo also contributed to the formation of what is now called Newton’s first law of motion. Newton made use of the work of his predecessors, which enabled him to develop laws of motion, discover the law of gravity, invent calculus, and make great contributions to the theories of light and color. It is amazing that many of these developments were made with Newton working alone, without the benefit of the usual interactions that take place among scientists today.

It was not until the advent of modern physics early in the 20th century that it was discovered that Newton’s laws of motion produce a good approximation to motion only when the objects are moving at speeds much, much less than the speed of light and when those objects are larger than the

size of most molecules (about 10^{-9} m in diameter). These constraints define the realm of classical mechanics, as discussed in [Introduction to the Nature of Science and Physics](#). At the beginning of the 20th century, Albert Einstein (1879–1955) developed the theory of relativity and, along with many other scientists, developed quantum theory. This theory does not have the constraints present in classical physics. All of the situations we consider in this chapter, and all those preceding the introduction of relativity in [Special Relativity](#), are in the realm of classical physics.

Note:

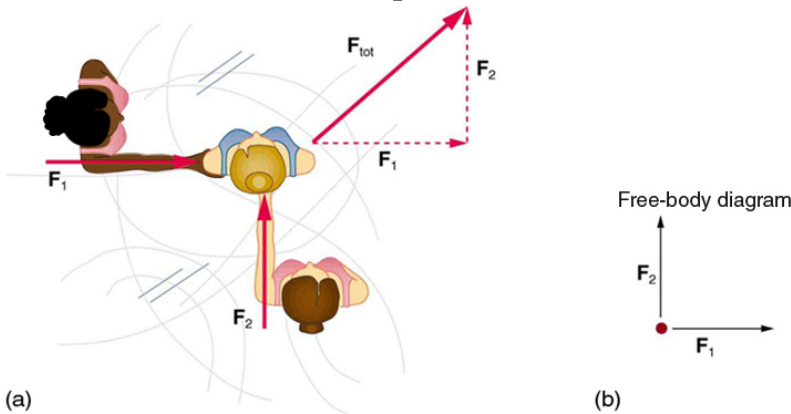
Making Connections: Past and Present Philosophy

The importance of observation and the concept of *cause and effect* were not always so entrenched in human thinking. This realization was a part of the evolution of modern physics from natural philosophy. The achievements of Galileo, Newton, Einstein, and others were key milestones in the history of scientific thought. Most of the scientific theories that are described in this book descended from the work of these scientists.

Development of Force Concept

- Understand the definition of force.

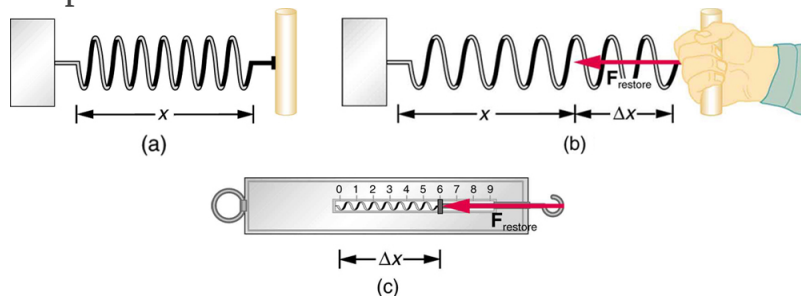
Dynamics is the study of the forces that cause objects and systems to move. To understand this, we need a working definition of force. Our intuitive definition of **force**—that is, a push or a pull—is a good place to start. We know that a push or pull has both magnitude and direction (therefore, it is a vector quantity) and can vary considerably in each regard. For example, a cannon exerts a strong force on a cannonball that is launched into the air. In contrast, Earth exerts only a tiny downward pull on a flea. Our everyday experiences also give us a good idea of how multiple forces add. If two people push in different directions on a third person, as illustrated in [\[link\]](#), we might expect the total force to be in the direction shown. Since force is a vector, it adds just like other vectors, as illustrated in [\[link\]\(a\)](#) for two ice skaters. Forces, like other vectors, are represented by arrows and can be added using the familiar head-to-tail method or by trigonometric methods. These ideas were developed in [Two-Dimensional Kinematics](#).



Part (a) shows an overhead view of two ice skaters pushing on a third. Forces are vectors and add like other vectors, so the total force on the third skater is in the direction shown. In part (b), we see a free-body diagram representing the forces acting on the third skater.

[\[link\]](#)(b) is our first example of a **free-body diagram**, which is a technique used to illustrate all the **external forces** acting on a body. The body is represented by a single isolated point (or free body), and only those forces acting *on* the body from the outside (external forces) are shown. (These forces are the only ones shown, because only external forces acting on the body affect its motion. We can ignore any internal forces within the body.) Free-body diagrams are very useful in analyzing forces acting on a system and are employed extensively in the study and application of Newton's laws of motion.

A more quantitative definition of force can be based on some standard force, just as distance is measured in units relative to a standard distance. One possibility is to stretch a spring a certain fixed distance, as illustrated in [\[link\]](#), and use the force it exerts to pull itself back to its relaxed shape—called a *restoring force*—as a standard. The magnitude of all other forces can be stated as multiples of this standard unit of force. Many other possibilities exist for standard forces. (One that we will encounter in [Magnetism](#) is the magnetic force between two wires carrying electric current.) Some alternative definitions of force will be given later in this chapter.



The force exerted by a stretched spring can be used as a standard unit of force. (a) This spring has a length x when undistorted. (b) When stretched a distance Δx , the spring exerts a restoring force, $\mathbf{F}_{\text{restore}}$, which is reproducible. (c) A spring scale is one device that uses a spring to measure force. The force $\mathbf{F}_{\text{restore}}$ is exerted on whatever is attached to the hook. Here $\mathbf{F}_{\text{restore}}$ has a

magnitude of 6 units in the force standard being employed.

Note:

Take-Home Experiment: Force Standards

To investigate force standards and cause and effect, get two identical rubber bands. Hang one rubber band vertically on a hook. Find a small household item that could be attached to the rubber band using a paper clip, and use this item as a weight to investigate the stretch of the rubber band. Measure the amount of stretch produced in the rubber band with one, two, and four of these (identical) items suspended from the rubber band. What is the relationship between the number of items and the amount of stretch? How large a stretch would you expect for the same number of items suspended from two rubber bands? What happens to the amount of stretch of the rubber band (with the weights attached) if the weights are also pushed to the side with a pencil?

Section Summary

- **Dynamics** is the study of how forces affect the motion of objects.
- **Force** is a push or pull that can be defined in terms of various standards, and it is a vector having both magnitude and direction.
- **External forces** are any outside forces that act on a body. A **free-body diagram** is a drawing of all external forces acting on a body.

Conceptual Questions

Exercise:

Problem:

Propose a force standard different from the example of a stretched spring discussed in the text. Your standard must be capable of producing the same force repeatedly.

Exercise:**Problem:**

What properties do forces have that allow us to classify them as vectors?

Glossary

dynamics

the study of how forces affect the motion of objects and systems

external force

a force acting on an object or system that originates outside of the object or system

free-body diagram

a sketch showing all of the external forces acting on an object or system; the system is represented by a dot, and the forces are represented by vectors extending outward from the dot

force

a push or pull on an object with a specific magnitude and direction; can be represented by vectors; can be expressed as a multiple of a standard force

Newton's First Law of Motion: Inertia

- Define mass and inertia.
- Understand Newton's first law of motion.

Experience suggests that an object at rest will remain at rest if left alone, and that an object in motion tends to slow down and stop unless some effort is made to keep it moving. What **Newton's first law of motion** states, however, is the following:

Note:

Newton's First Law of Motion

A body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force.

Note the repeated use of the verb “remains.” We can think of this law as preserving the status quo of motion.

Rather than contradicting our experience, **Newton's first law of motion** states that there must be a *cause* (which is a net external force) *for there to be any change in velocity (either a change in magnitude or direction)*. We will define *net external force* in the next section. An object sliding across a table or floor slows down due to the net force of friction acting on the object. If friction disappeared, would the object still slow down?

The idea of cause and effect is crucial in accurately describing what happens in various situations. For example, consider what happens to an object sliding along a rough horizontal surface. The object quickly grinds to a halt. If we spray the surface with talcum powder to make the surface smoother, the object slides farther. If we make the surface even smoother by rubbing lubricating oil on it, the object slides farther yet. Extrapolating to a frictionless surface, we can imagine the object sliding in a straight line indefinitely. Friction is thus the *cause* of the slowing (consistent with Newton's first law). The object would not slow down at all if friction were

completely eliminated. Consider an air hockey table. When the air is turned off, the puck slides only a short distance before friction slows it to a stop. However, when the air is turned on, it creates a nearly frictionless surface, and the puck glides long distances without slowing down. Additionally, if we know enough about the friction, we can accurately predict how quickly the object will slow down. Friction is an external force.

Newton's first law is completely general and can be applied to anything from an object sliding on a table to a satellite in orbit to blood pumped from the heart. Experiments have thoroughly verified that any change in velocity (speed or direction) must be caused by an external force. The idea of *generally applicable or universal laws* is important not only here—it is a basic feature of all laws of physics. Identifying these laws is like recognizing patterns in nature from which further patterns can be discovered. The genius of Galileo, who first developed the idea for the first law, and Newton, who clarified it, was to ask the fundamental question, “What is the cause?” Thinking in terms of cause and effect is a worldview fundamentally different from the typical ancient Greek approach when questions such as “Why does a tiger have stripes?” would have been answered in Aristotelian fashion, “That is the nature of the beast.” True perhaps, but not a useful insight.

Mass

The property of a body to remain at rest or to remain in motion with constant velocity is called **inertia**. Newton's first law is often called the **law of inertia**. As we know from experience, some objects have more inertia than others. It is obviously more difficult to change the motion of a large boulder than that of a basketball, for example. The inertia of an object is measured by its **mass**. Roughly speaking, mass is a measure of the amount of “stuff” (or matter) in something. The quantity or amount of matter in an object is determined by the numbers of atoms and molecules of various types it contains. Unlike weight, mass does not vary with location. The mass of an object is the same on Earth, in orbit, or on the surface of the Moon. In practice, it is very difficult to count and identify all of the atoms and molecules in an object, so masses are not often determined in this

manner. Operationally, the masses of objects are determined by comparison with the standard kilogram.

Exercise:

Check Your Understanding

Problem:

Which has more mass: a kilogram of cotton balls or a kilogram of gold?

Solution:

Answer

They are equal. A kilogram of one substance is equal in mass to a kilogram of another substance. The quantities that might differ between them are volume and density.

Section Summary

- **Newton's first law of motion** states that a body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force. This is also known as the **law of inertia**.
- **Inertia** is the tendency of an object to remain at rest or remain in motion. Inertia is related to an object's mass.
- **Mass** is the quantity of matter in a substance.

Conceptual Questions

Exercise:

Problem: How are inertia and mass related?

Exercise:

Problem:

What is the relationship between weight and mass? Which is an intrinsic, unchanging property of a body?

Glossary

inertia

the tendency of an object to remain at rest or remain in motion

law of inertia

see Newton's first law of motion

mass

the quantity of matter in a substance; measured in kilograms

Newton's first law of motion

a body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force; also known as the law of inertia

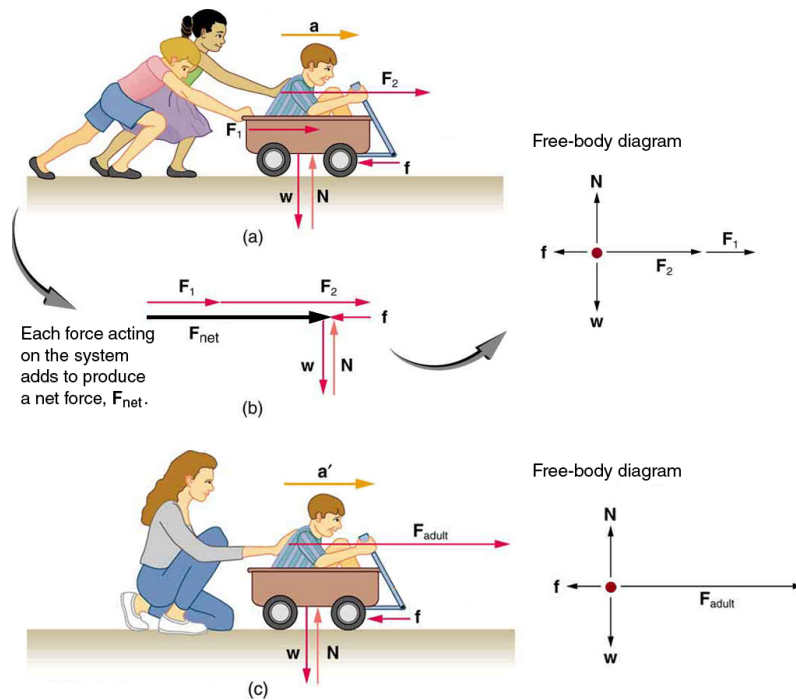
Newton's Second Law of Motion: Concept of a System

- Define net force, external force, and system.
- Understand Newton's second law of motion.
- Apply Newton's second law to determine the weight of an object.

Newton's second law of motion is closely related to Newton's first law of motion. It mathematically states the cause and effect relationship between force and changes in motion. Newton's second law of motion is more quantitative and is used extensively to calculate what happens in situations involving a force. Before we can write down Newton's second law as a simple equation giving the exact relationship of force, mass, and acceleration, we need to sharpen some ideas that have already been mentioned.

First, what do we mean by a change in motion? The answer is that a change in motion is equivalent to a change in velocity. A change in velocity means, by definition, that there is an **acceleration**. Newton's first law says that a net external force causes a change in motion; thus, we see that a *net external force causes acceleration*.

Another question immediately arises. What do we mean by an external force? An intuitive notion of external is correct—an **external force** acts from outside the **system** of interest. For example, in [\[link\]](#)(a) the system of interest is the wagon plus the child in it. The two forces exerted by the other children are external forces. An internal force acts between elements of the system. Again looking at [\[link\]](#)(a), the force the child in the wagon exerts to hang onto the wagon is an internal force between elements of the system of interest. Only external forces affect the motion of a system, according to Newton's first law. (The internal forces actually cancel, as we shall see in the next section.) *You must define the boundaries of the system before you can determine which forces are external.* Sometimes the system is obvious, whereas other times identifying the boundaries of a system is more subtle. The concept of a system is fundamental to many areas of physics, as is the correct application of Newton's laws. This concept will be revisited many times on our journey through physics.



Different forces exerted on the same mass produce different accelerations. (a) Two children push a wagon with a child in it. Arrows representing all external forces are shown. The system of interest is the wagon and its rider. The weight w of the system and the support of the ground N are also shown for completeness and are assumed to cancel. The vector f represents the friction acting on the wagon, and it acts to the left, opposing the motion of the wagon. (b) All of the external forces acting on the system add together to produce a net force, F_{net} . The free-body diagram shows all of the forces acting on the system of interest. The dot represents the center of mass of the system. Each force vector extends from this dot. Because there are two forces acting to the right, we draw the vectors collinearly. (c) A larger net external force produces a larger

acceleration ($\mathbf{a}' > \mathbf{a}$) when an adult pushes the child.

Now, it seems reasonable that acceleration should be directly proportional to and in the same direction as the net (total) external force acting on a system. This assumption has been verified experimentally and is illustrated in [\[link\]](#). In part (a), a smaller force causes a smaller acceleration than the larger force illustrated in part (c). For completeness, the vertical forces are also shown; they are assumed to cancel since there is no acceleration in the vertical direction. The vertical forces are the weight \mathbf{w} and the support of the ground \mathbf{N} , and the horizontal force \mathbf{f} represents the force of friction. These will be discussed in more detail in later sections. For now, we will define **friction** as a force that opposes the motion past each other of objects that are touching. [\[link\]](#)(b) shows how vectors representing the external forces add together to produce a net force, \mathbf{F}_{net} .

To obtain an equation for Newton's second law, we first write the relationship of acceleration and net external force as the proportionality **Equation:**

$$\mathbf{a} \propto \mathbf{F}_{\text{net}},$$

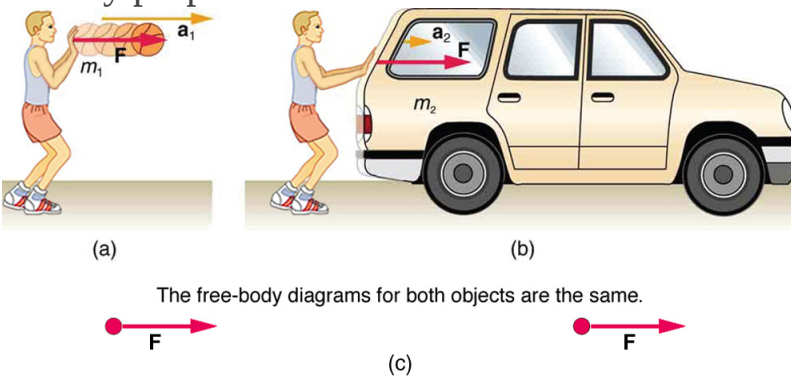
where the symbol \propto means “proportional to,” and \mathbf{F}_{net} is the **net external force**. (The net external force is the vector sum of all external forces and can be determined graphically, using the head-to-tail method, or analytically, using components. The techniques are the same as for the addition of other vectors, and are covered in [Two-Dimensional Kinematics](#).) This proportionality states what we have said in words—*acceleration is directly proportional to the net external force*. Once the system of interest is chosen, it is important to identify the external forces and ignore the internal ones. It is a tremendous simplification not to have to consider the numerous internal forces acting between objects within the system, such as muscular forces within the child's body, let alone the myriad of forces between atoms in the objects, but by doing so, we can easily solve some very complex problems with only minimal error due to our simplification

Now, it also seems reasonable that acceleration should be inversely proportional to the mass of the system. In other words, the larger the mass (the inertia), the smaller the acceleration produced by a given force. And indeed, as illustrated in [\[link\]](#), the same net external force applied to a car produces a much smaller acceleration than when applied to a basketball. The proportionality is written as

Equation:

$$\mathbf{a} \propto \frac{1}{m}$$

where m is the mass of the system. Experiments have shown that acceleration is exactly inversely proportional to mass, just as it is exactly linearly proportional to the net external force.



The same force exerted on systems of different masses produces different accelerations. (a) A basketball player pushes on a basketball to make a pass. (The effect of gravity on the ball is ignored.) (b) The same player exerts an identical force on a stalled SUV and produces a far smaller acceleration (even if friction is negligible). (c) The free-body diagrams are identical, permitting direct comparison of the two situations. A series of patterns for the free-body diagram will emerge as you do more problems.

It has been found that the acceleration of an object depends *only* on the net external force and the mass of the object. Combining the two proportionalities just given yields Newton's second law of motion.

Note:

Newton's Second Law of Motion

The acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to its mass.

In equation form, Newton's second law of motion is

Equation:

$$\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}.$$

This is often written in the more familiar form

Equation:

$$\mathbf{F}_{\text{net}} = m\mathbf{a}.$$

When only the magnitude of force and acceleration are considered, this equation is simply

Equation:

$$F_{\text{net}} = ma.$$

Although these last two equations are really the same, the first gives more insight into what Newton's second law means. The law is a *cause and effect relationship* among three quantities that is not simply based on their definitions. The validity of the second law is completely based on experimental verification.

Units of Force

$\mathbf{F}_{\text{net}} = m\mathbf{a}$ is used to define the units of force in terms of the three basic units for mass, length, and time. The SI unit of force is called the **newton** (abbreviated N) and is the force needed to accelerate a 1-kg system at the rate of 1m/s^2 . That is, since $\mathbf{F}_{\text{net}} = m\mathbf{a}$,

Equation:

$$1 \text{ N} = 1 \text{ kg} \cdot \text{m/s}^2.$$

While almost the entire world uses the newton for the unit of force, in the United States the most familiar unit of force is the pound (lb), where $1 \text{ N} = 0.225 \text{ lb}$.

Weight and the Gravitational Force

When an object is dropped, it accelerates toward the center of Earth. Newton's second law states that a net force on an object is responsible for its acceleration. If air resistance is negligible, the net force on a falling object is the gravitational force, commonly called its **weight \mathbf{w}** . Weight can be denoted as a vector \mathbf{w} because it has a direction; *down* is, by definition, the direction of gravity, and hence weight is a downward force. The magnitude of weight is denoted as w . Galileo was instrumental in showing that, in the absence of air resistance, all objects fall with the same acceleration g . Using Galileo's result and Newton's second law, we can derive an equation for weight.

Consider an object with mass m falling downward toward Earth. It experiences only the downward force of gravity, which has magnitude w . Newton's second law states that the magnitude of the net external force on an object is $F_{\text{net}} = ma$.

Since the object experiences only the downward force of gravity, $F_{\text{net}} = w$. We know that the acceleration of an object due to gravity is g , or $a = g$. Substituting these into Newton's second law gives

Note:**Weight**

This is the equation for *weight*—the gravitational force on a mass m :

Equation:

$$w = mg.$$

Since $g = 9.80 \text{ m/s}^2$ on Earth, the weight of a 1.0 kg object on Earth is 9.8 N, as we see:

Equation:

$$w = mg = (1.0 \text{ kg})(9.80 \text{ m/s}^2) = 9.8 \text{ N}.$$

Recall that g can take a positive or negative value, depending on the positive direction in the coordinate system. Be sure to take this into consideration when solving problems with weight.

When the net external force on an object is its weight, we say that it is in **free-fall**. That is, the only force acting on the object is the force of gravity. In the real world, when objects fall downward toward Earth, they are never truly in free-fall because there is always some upward force from the air acting on the object.

The acceleration due to gravity g varies slightly over the surface of Earth, so that the weight of an object depends on location and is not an intrinsic property of the object. Weight varies dramatically if one leaves Earth's surface. On the Moon, for example, the acceleration due to gravity is only 1.67 m/s^2 . A 1.0-kg mass thus has a weight of 9.8 N on Earth and only about 1.7 N on the Moon.

The broadest definition of weight in this sense is that *the weight of an object is the gravitational force on it from the nearest large body*, such as Earth, the Moon, the Sun, and so on. This is the most common and useful definition of weight in physics. It differs dramatically, however, from the definition of weight used by NASA and the popular media in relation to space travel and exploration. When they speak of “weightlessness” and

“microgravity,” they are really referring to the phenomenon we call “free-fall” in physics. We shall use the above definition of weight, and we will make careful distinctions between free-fall and actual weightlessness.

It is important to be aware that weight and mass are very different physical quantities, although they are closely related. Mass is the quantity of matter (how much “stuff”) and does not vary in classical physics, whereas weight is the gravitational force and does vary depending on gravity. It is tempting to equate the two, since most of our examples take place on Earth, where the weight of an object only varies a little with the location of the object. Furthermore, the terms *mass* and *weight* are used interchangeably in everyday language; for example, our medical records often show our “weight” in kilograms, but never in the correct units of newtons.

Note:**Common Misconceptions: Mass vs. Weight**

Mass and weight are often used interchangeably in everyday language. However, in science, these terms are distinctly different from one another. Mass is a measure of how much matter is in an object. The typical measure of mass is the kilogram (or the “slug” in English units). Weight, on the other hand, is a measure of the force of gravity acting on an object. Weight is equal to the mass of an object (m) multiplied by the acceleration due to gravity (g). Like any other force, weight is measured in terms of newtons (or pounds in English units).

Assuming the mass of an object is kept intact, it will remain the same, regardless of its location. However, because weight depends on the acceleration due to gravity, the weight of an object *can change* when the object enters into a region with stronger or weaker gravity. For example, the acceleration due to gravity on the Moon is 1.67 m/s^2 (which is much less than the acceleration due to gravity on Earth, 9.80 m/s^2). If you measured your weight on Earth and then measured your weight on the Moon, you would find that you “weigh” much less, even though you do not look any skinnier. This is because the force of gravity is weaker on the Moon. In fact, when people say that they are “losing weight,” they really

mean that they are losing “mass” (which in turn causes them to weigh less).

Note:

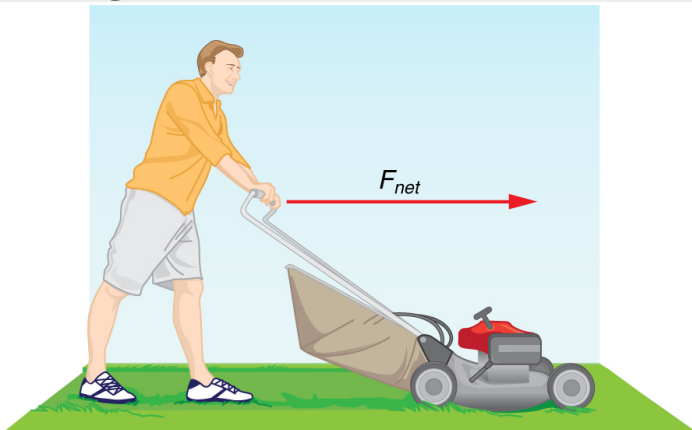
Take-Home Experiment: Mass and Weight

What do bathroom scales measure? When you stand on a bathroom scale, what happens to the scale? It depresses slightly. The scale contains springs that compress in proportion to your weight—similar to rubber bands expanding when pulled. The springs provide a measure of your weight (for an object which is not accelerating). This is a force in newtons (or pounds). In most countries, the measurement is divided by 9.80 to give a reading in mass units of kilograms. The scale measures weight but is calibrated to provide information about mass. While standing on a bathroom scale, push down on a table next to you. What happens to the reading? Why? Would your scale measure the same “mass” on Earth as on the Moon?

Example:

What Acceleration Can a Person Produce when Pushing a Lawn Mower?

Suppose that the net external force (push minus friction) exerted on a lawn mower is 51 N (about 11 lb) parallel to the ground. The mass of the mower is 24 kg. What is its acceleration?



The net force on a lawn mower is 51

N to the right. At what rate does the lawn mower accelerate to the right?

Strategy

Since \mathbf{F}_{net} and m are given, the acceleration can be calculated directly from Newton's second law as stated in $\mathbf{F}_{\text{net}} = m\mathbf{a}$.

Solution

The magnitude of the acceleration a is $a = \frac{F_{\text{net}}}{m}$. Entering known values gives

Equation:

$$a = \frac{51 \text{ N}}{24 \text{ kg}}$$

Substituting the units $\text{kg} \cdot \text{m}/\text{s}^2$ for N yields

Equation:

$$a = \frac{51 \text{ kg} \cdot \text{m}/\text{s}^2}{24 \text{ kg}} = 2.1 \text{ m}/\text{s}^2.$$

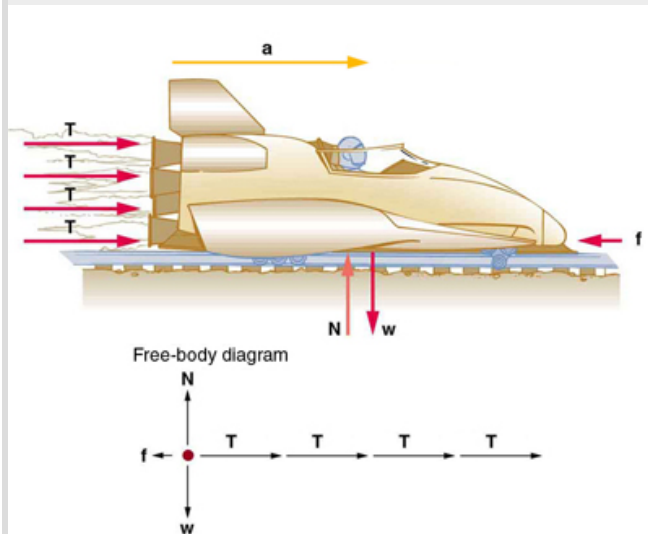
Discussion

The direction of the acceleration is the same direction as that of the net force, which is parallel to the ground. There is no information given in this example about the individual external forces acting on the system, but we can say something about their relative magnitudes. For example, the force exerted by the person pushing the mower must be greater than the friction opposing the motion (since we know the mower moves forward), and the vertical forces must cancel if there is to be no acceleration in the vertical direction (the mower is moving only horizontally). The acceleration found is small enough to be reasonable for a person pushing a mower. Such an effort would not last too long because the person's top speed would soon be reached.

Example:

What Rocket Thrust Accelerates This Sled?

Prior to manned space flights, rocket sleds were used to test aircraft, missile equipment, and physiological effects on human subjects at high speeds. They consisted of a platform that was mounted on one or two rails and propelled by several rockets. Calculate the magnitude of force exerted by each rocket, called its thrust \mathbf{T} , for the four-rocket propulsion system shown in [\[link\]](#). The sled's initial acceleration is 49 m/s^2 , the mass of the system is 2100 kg , and the force of friction opposing the motion is known to be 650 N .



A sled experiences a rocket thrust that accelerates it to the right. Each rocket creates an identical thrust \mathbf{T} . As in other situations where there is only horizontal acceleration, the vertical forces cancel. The ground exerts an upward force \mathbf{N} on the system that is equal in magnitude and opposite in direction to its weight, \mathbf{w} . The system here is the sled, its rockets, and rider, so none of the forces *between* these objects are considered. The arrow representing friction (\mathbf{f}) is drawn larger than scale.

Strategy

Although there are forces acting vertically and horizontally, we assume the vertical forces cancel since there is no vertical acceleration. This leaves us with only horizontal forces and a simpler one-dimensional problem.

Directions are indicated with plus or minus signs, with right taken as the positive direction. See the free-body diagram in the figure.

Solution

Since acceleration, mass, and the force of friction are given, we start with Newton's second law and look for ways to find the thrust of the engines.

Since we have defined the direction of the force and acceleration as acting "to the right," we need to consider only the magnitudes of these quantities in the calculations. Hence we begin with

Equation:

$$F_{\text{net}} = ma,$$

where F_{net} is the net force along the horizontal direction. We can see from [\[link\]](#) that the engine thrusts add, while friction opposes the thrust. In equation form, the net external force is

Equation:

$$F_{\text{net}} = 4T - f.$$

Substituting this into Newton's second law gives

Equation:

$$F_{\text{net}} = ma = 4T - f.$$

Using a little algebra, we solve for the total thrust $4T$:

Equation:

$$4T = ma + f.$$

Substituting known values yields

Equation:

$$4T = ma + f = (2100 \text{ kg})(49 \text{ m/s}^2) + 650 \text{ N}.$$

So the total thrust is

Equation:

$$4T = 1.0 \times 10^5 \text{ N},$$

and the individual thrusts are

Equation:

$$T = \frac{1.0 \times 10^5 \text{ N}}{4} = 2.6 \times 10^4 \text{ N}.$$

Discussion

The numbers are quite large, so the result might surprise you. Experiments such as this were performed in the early 1960s to test the limits of human endurance and the setup designed to protect human subjects in jet fighter emergency ejections. Speeds of 1000 km/h were obtained, with accelerations of 45 *g*'s. (Recall that *g*, the acceleration due to gravity, is 9.80 m/s². When we say that an acceleration is 45 *g*'s, it is 45 × 9.80 m/s², which is approximately 440 m/s².) While living subjects are not used any more, land speeds of 10,000 km/h have been obtained with rocket sleds. In this example, as in the preceding one, the system of interest is obvious. We will see in later examples that choosing the system of interest is crucial—and the choice is not always obvious.

Newton's second law of motion is more than a definition; it is a relationship among acceleration, force, and mass. It can help us make predictions. Each of those physical quantities can be defined independently, so the second law tells us something basic and universal about nature. The next section introduces the third and final law of motion.

Section Summary

- Acceleration, **a**, is defined as a change in velocity, meaning a change in its magnitude or direction, or both.
- An external force is one acting on a system from outside the system, as opposed to internal forces, which act between components within the

system.

- Newton's second law of motion states that the acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to its mass.
- In equation form, Newton's second law of motion is $\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}$.
- This is often written in the more familiar form: $\mathbf{F}_{\text{net}} = m\mathbf{a}$.
- The weight \mathbf{w} of an object is defined as the force of gravity acting on an object of mass m . The object experiences an acceleration due to gravity \mathbf{g} :

Equation:

$$\mathbf{w} = m\mathbf{g}.$$

- If the only force acting on an object is due to gravity, the object is in free fall.
- Friction is a force that opposes the motion past each other of objects that are touching.

Conceptual Questions

Exercise:

Problem:

Which statement is correct? (a) Net force causes motion. (b) Net force causes change in motion. Explain your answer and give an example.

Exercise:

Problem:

Why can we neglect forces such as those holding a body together when we apply Newton's second law of motion?

Exercise:

Problem:

Explain how the choice of the “system of interest” affects which forces must be considered when applying Newton’s second law of motion.

Exercise:

Problem:

Describe a situation in which the net external force on a system is not zero, yet its speed remains constant.

Exercise:

Problem:

A system can have a nonzero velocity while the net external force on it is zero. Describe such a situation.

Exercise:

Problem:

A rock is thrown straight up. What is the net external force acting on the rock when it is at the top of its trajectory?

Exercise:

Problem:

(a) Give an example of different net external forces acting on the same system to produce different accelerations. (b) Give an example of the same net external force acting on systems of different masses, producing different accelerations. (c) What law accurately describes both effects? State it in words and as an equation.

Exercise:

Problem:

If the acceleration of a system is zero, are no external forces acting on it? What about internal forces? Explain your answers.

Exercise:

Problem:

If a constant, nonzero force is applied to an object, what can you say about the velocity and acceleration of the object?

Exercise:**Problem:**

The gravitational force on the basketball in [\[link\]](#) is ignored. When gravity *is* taken into account, what is the direction of the net external force on the basketball—above horizontal, below horizontal, or still horizontal?

Problem Exercises

You may assume data taken from illustrations is accurate to three digits.

Exercise:**Problem:**

A 63.0-kg sprinter starts a race with an acceleration of 4.20 m/s^2 . What is the net external force on him?

Solution:

265 N

Exercise:**Problem:**

If the sprinter from the previous problem accelerates at that rate for 20 m, and then maintains that velocity for the remainder of the 100-m dash, what will be his time for the race?

Exercise:

Problem:

A cleaner pushes a 4.50-kg laundry cart in such a way that the net external force on it is 60.0 N. Calculate the magnitude of its acceleration.

Solution:

$$13.3 \text{ m/s}^2$$

Exercise:**Problem:**

Since astronauts in orbit are apparently weightless, a clever method of measuring their masses is needed to monitor their mass gains or losses to adjust diets. One way to do this is to exert a known force on an astronaut and measure the acceleration produced. Suppose a net external force of 50.0 N is exerted and the astronaut's acceleration is measured to be 0.893 m/s^2 . (a) Calculate her mass. (b) By exerting a force on the astronaut, the vehicle in which they orbit experiences an equal and opposite force. Discuss how this would affect the measurement of the astronaut's acceleration. Propose a method in which recoil of the vehicle is avoided.

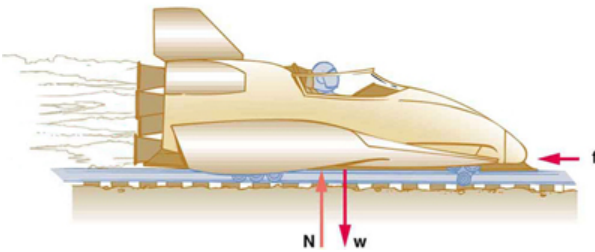
Exercise:**Problem:**

In [\[link\]](#), the net external force on the 24-kg mower is stated to be 51 N. If the force of friction opposing the motion is 24 N, what force F (in newtons) is the person exerting on the mower? Suppose the mower is moving at 1.5 m/s when the force F is removed. How far will the mower go before stopping?

Exercise:

Problem:

The same rocket sled drawn in [\[link\]](#) is decelerated at a rate of 196 m/s^2 . What force is necessary to produce this deceleration? Assume that the rockets are off. The mass of the system is 2100 kg.

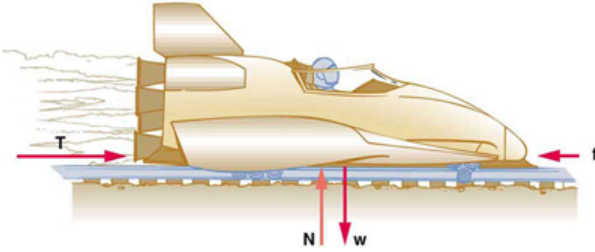
**Exercise:****Problem:**

(a) If the rocket sled shown in [\[link\]](#) starts with only one rocket burning, what is the magnitude of its acceleration? Assume that the mass of the system is 2100 kg, the thrust T is $2.4 \times 10^4 \text{ N}$, and the force of friction opposing the motion is known to be 650 N. (b) Why is the acceleration not one-fourth of what it is with all rockets burning?

Solution:

(a) 12 m/s^2 .

(b) The acceleration is not one-fourth of what it was with all rockets burning because the frictional force is still as large as it was with all rockets burning.



Exercise:

Problem:

What is the deceleration of the rocket sled if it comes to rest in 1.1 s from a speed of 1000 km/h? (Such deceleration caused one test subject to black out and have temporary blindness.)

Exercise:

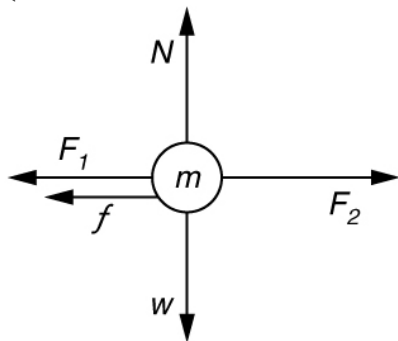
Problem:

Suppose two children push horizontally, but in exactly opposite directions, on a third child in a wagon. The first child exerts a force of 75.0 N, the second a force of 90.0 N, friction is 12.0 N, and the mass of the third child plus wagon is 23.0 kg. (a) What is the system of interest if the acceleration of the child in the wagon is to be calculated? (b) Draw a free-body diagram, including all forces acting on the system. (c) Calculate the acceleration. (d) What would the acceleration be if friction were 15.0 N?

Solution:

(a) The system is the child in the wagon plus the wagon.

(b)



(c) $a = 0.130 \text{ m/s}^2$ in the direction of the second child's push.

(d) $a = 0.00 \text{ m/s}^2$

Exercise:

Problem:

A powerful motorcycle can produce an acceleration of 3.50 m/s^2 while traveling at 90.0 km/h . At that speed the forces resisting motion, including friction and air resistance, total 400 N . (Air resistance is analogous to air friction. It always opposes the motion of an object.) What is the magnitude of the force the motorcycle exerts backward on the ground to produce its acceleration if the mass of the motorcycle with rider is 245 kg ?

Exercise:

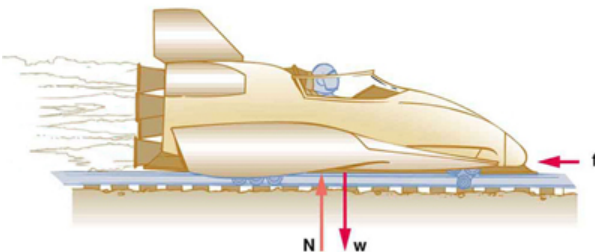
Problem:

The rocket sled shown in [\[link\]](#) accelerates at a rate of 49.0 m/s^2 . Its passenger has a mass of 75.0 kg . (a) Calculate the horizontal component of the force the seat exerts against his body. Compare this with his weight by using a ratio. (b) Calculate the direction and magnitude of the total force the seat exerts against his body.

Solution:

(a) $3.68 \times 10^3 \text{ N}$. This force is 5.00 times greater than his weight.

(b) 3750 N ; 11.3° above horizontal



Exercise:**Problem:**

Repeat the previous problem for the situation in which the rocket sled decelerates at a rate of 201 m/s^2 . In this problem, the forces are exerted by the seat and restraining belts.

Exercise:**Problem:**

The weight of an astronaut plus his space suit on the Moon is only 250 N. How much do they weigh on Earth? What is the mass on the Moon? On Earth?

Solution:

$1.5 \times 10^3 \text{ N}$, 150 kg, 150 kg

Exercise:**Problem:**

Suppose the mass of a fully loaded module in which astronauts take off from the Moon is 10,000 kg. The thrust of its engines is 30,000 N. (a) Calculate its the magnitude of acceleration in a vertical takeoff from the Moon. (b) Could it lift off from Earth? If not, why not? If it could, calculate the magnitude of its acceleration.

Glossary

acceleration

the rate at which an object's velocity changes over a period of time

free-fall

a situation in which the only force acting on an object is the force due to gravity

friction

a force past each other of objects that are touching; examples include rough surfaces and air resistance

net external force

the vector sum of all external forces acting on an object or system; causes a mass to accelerate

Newton's second law of motion

the net external force \mathbf{F}_{net} on an object with mass m is proportional to and in the same direction as the acceleration of the object, \mathbf{a} , and inversely proportional to the mass; defined mathematically as

$$\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}$$

system

defined by the boundaries of an object or collection of objects being observed; all forces originating from outside of the system are considered external forces

weight

the force \mathbf{w} due to gravity acting on an object of mass m ; defined mathematically as: $\mathbf{w} = m\mathbf{g}$, where \mathbf{g} is the magnitude and direction of the acceleration due to gravity

Newton's Third Law of Motion: Symmetry in Forces

- Understand Newton's third law of motion.
- Apply Newton's third law to define systems and solve problems of motion.

There is a passage in the musical *Man of la Mancha* that relates to Newton's third law of motion. Sancho, in describing a fight with his wife to Don Quixote, says, "Of course I hit her back, Your Grace, but she's a lot harder than me and you know what they say, 'Whether the stone hits the pitcher or the pitcher hits the stone, it's going to be bad for the pitcher.'" This is exactly what happens whenever one body exerts a force on another—the first also experiences a force (equal in magnitude and opposite in direction). Numerous common experiences, such as stubbing a toe or throwing a ball, confirm this. It is precisely stated in **Newton's third law of motion**.

Note:

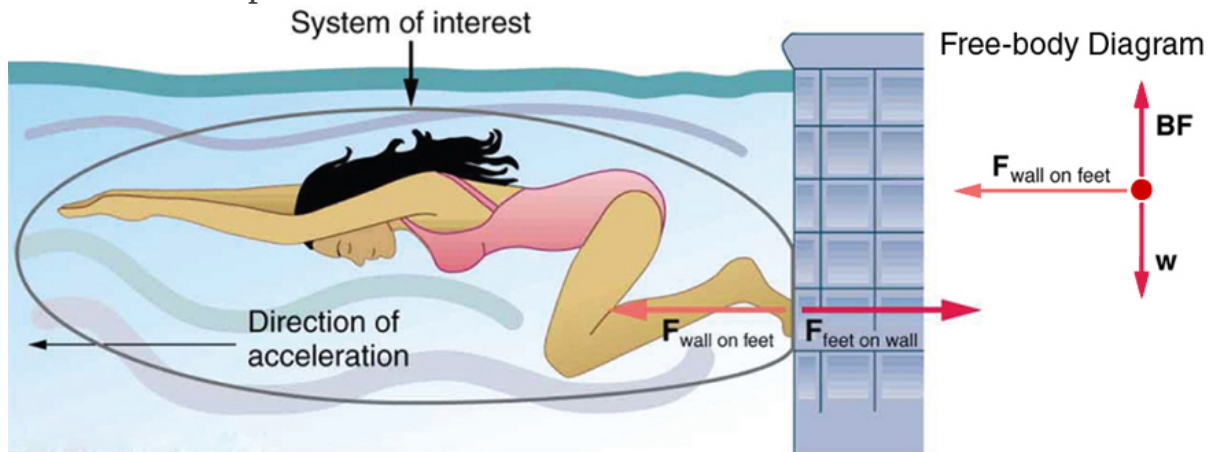
Newton's Third Law of Motion

Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts.

This law represents a certain *symmetry in nature*: Forces always occur in pairs, and one body cannot exert a force on another without experiencing a force itself. We sometimes refer to this law loosely as "action-reaction," where the force exerted is the action and the force experienced as a consequence is the reaction. Newton's third law has practical uses in analyzing the origin of forces and understanding which forces are external to a system.

We can readily see Newton's third law at work by taking a look at how people move about. Consider a swimmer pushing off from the side of a pool, as illustrated in [\[link\]](#). She pushes against the pool wall with her feet

and accelerates in the direction *opposite* to that of her push. The wall has exerted an equal and opposite force back on the swimmer. You might think that two equal and opposite forces would cancel, but they do not *because they act on different systems*. In this case, there are two systems that we could investigate: the swimmer or the wall. If we select the swimmer to be the system of interest, as in the figure, then $\mathbf{F}_{\text{wall on feet}}$ is an external force on this system and affects its motion. The swimmer moves in the direction of $\mathbf{F}_{\text{wall on feet}}$. In contrast, the force $\mathbf{F}_{\text{feet on wall}}$ acts on the wall and not on our system of interest. Thus $\mathbf{F}_{\text{feet on wall}}$ does not directly affect the motion of the system and does not cancel $\mathbf{F}_{\text{wall on feet}}$. Note that the swimmer pushes in the direction opposite to that in which she wishes to move. The reaction to her push is thus in the desired direction.

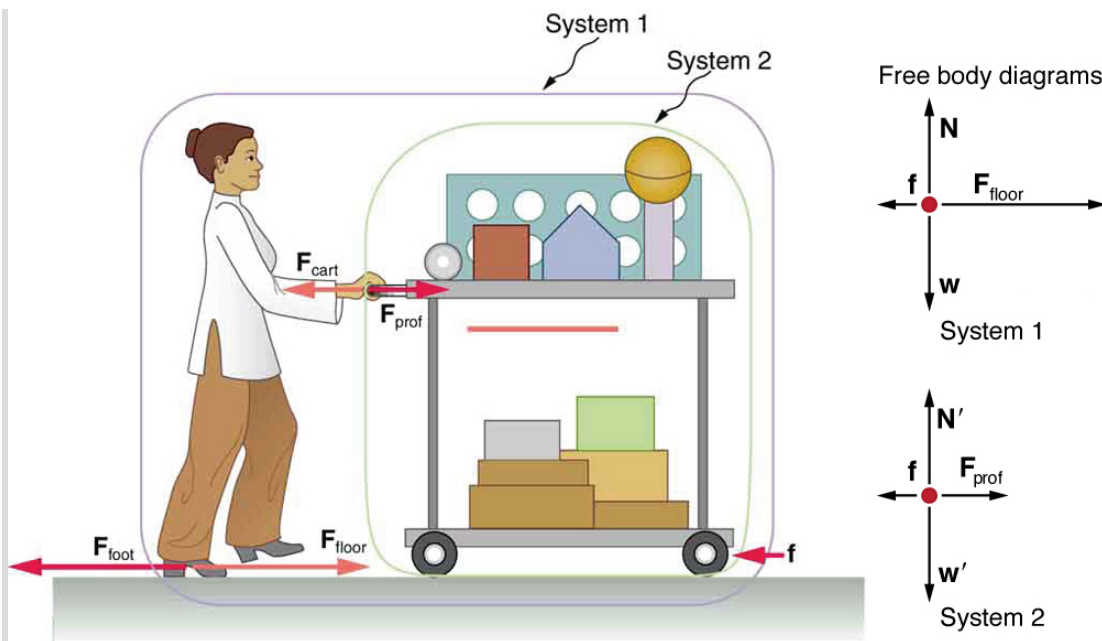


When the swimmer exerts a force $\mathbf{F}_{\text{feet on wall}}$ on the wall, she accelerates in the direction opposite to that of her push. This means the net external force on her is in the direction opposite to $\mathbf{F}_{\text{feet on wall}}$. This opposition occurs because, in accordance with Newton's third law of motion, the wall exerts a force $\mathbf{F}_{\text{wall on feet}}$ on her, equal in magnitude but in the direction opposite to the one she exerts on it. The line around the swimmer indicates the system of interest. Note that $\mathbf{F}_{\text{feet on wall}}$ does not act on this system (the swimmer) and, thus, does not cancel $\mathbf{F}_{\text{wall on feet}}$. Thus the free-body diagram shows only $\mathbf{F}_{\text{wall on feet}}$, \mathbf{w} , the gravitational force, and \mathbf{BF} , the buoyant force of the water supporting the swimmer's weight. The vertical forces \mathbf{w} and \mathbf{BF} cancel since there is no vertical motion.

Other examples of Newton's third law are easy to find. As a professor paces in front of a whiteboard, she exerts a force backward on the floor. The floor exerts a reaction force forward on the professor that causes her to accelerate forward. Similarly, a car accelerates because the ground pushes forward on the drive wheels in reaction to the drive wheels pushing backward on the ground. You can see evidence of the wheels pushing backward when tires spin on a gravel road and throw rocks backward. In another example, rockets move forward by expelling gas backward at high velocity. This means the rocket exerts a large backward force on the gas in the rocket combustion chamber, and the gas therefore exerts a large reaction force forward on the rocket. This reaction force is called **thrust**. It is a common misconception that rockets propel themselves by pushing on the ground or on the air behind them. They actually work better in a vacuum, where they can more readily expel the exhaust gases. Helicopters similarly create lift by pushing air down, thereby experiencing an upward reaction force. Birds and airplanes also fly by exerting force on air in a direction opposite to that of whatever force they need. For example, the wings of a bird force air downward and backward in order to get lift and move forward. An octopus propels itself in the water by ejecting water through a funnel from its body, similar to a jet ski. In a situation similar to Sancho's, professional cage fighters experience reaction forces when they punch, sometimes breaking their hand by hitting an opponent's body.

Example:**Getting Up To Speed: Choosing the Correct System**

A physics professor pushes a cart of demonstration equipment to a lecture hall, as seen in [\[link\]](#). Her mass is 65.0 kg, the cart's is 12.0 kg, and the equipment's is 7.0 kg. Calculate the acceleration produced when the professor exerts a backward force of 150 N on the floor. All forces opposing the motion, such as friction on the cart's wheels and air resistance, total 24.0 N.



A professor pushes a cart of demonstration equipment. The lengths of the arrows are proportional to the magnitudes of the forces (except for f , since it is too small to draw to scale). Different questions are asked in each example; thus, the system of interest must be defined differently for each. System 1 is appropriate for this example, since it asks for the acceleration of the entire group of objects. Only F_{floor} and f are external forces acting on System 1 along the line of motion. All other forces either cancel or act on the outside world. System 2 is chosen for [\[link\]](#) so that F_{prof} will be an external force and enter into Newton's second law. Note that the free-body diagrams, which allow us to apply Newton's second law, vary with the system chosen.

Strategy

Since they accelerate as a unit, we define the system to be the professor, cart, and equipment. This is System 1 in [\[link\]](#). The professor pushes backward with a force F_{foot} of 150 N. According to Newton's third law, the floor exerts a forward reaction force F_{floor} of 150 N on System 1. Because all motion is horizontal, we can assume there is no net force in the vertical direction. The problem is therefore one-dimensional along the

horizontal direction. As noted, \mathbf{f} opposes the motion and is thus in the opposite direction of $\mathbf{F}_{\text{floor}}$. Note that we do not include the forces \mathbf{F}_{prof} or \mathbf{F}_{cart} because these are internal forces, and we do not include \mathbf{F}_{foot} because it acts on the floor, not on the system. There are no other significant forces acting on System 1. If the net external force can be found from all this information, we can use Newton's second law to find the acceleration as requested. See the free-body diagram in the figure.

Solution

Newton's second law is given by

Equation:

$$a = \frac{F_{\text{net}}}{m}.$$

The net external force on System 1 is deduced from [\[link\]](#) and the discussion above to be

Equation:

$$F_{\text{net}} = F_{\text{floor}} - f = 150 \text{ N} - 24.0 \text{ N} = 126 \text{ N}.$$

The mass of System 1 is

Equation:

$$m = (65.0 + 12.0 + 7.0) \text{ kg} = 84 \text{ kg}.$$

These values of F_{net} and m produce an acceleration of

Equation:

$$a = \frac{F_{\text{net}}}{m},$$
$$a = \frac{126 \text{ N}}{84 \text{ kg}} = 1.5 \text{ m/s}^2.$$

Discussion

None of the forces between components of System 1, such as between the professor's hands and the cart, contribute to the net external force because they are internal to System 1. Another way to look at this is to note that forces between components of a system cancel because they are equal in magnitude and opposite in direction. For example, the force exerted by the

professor on the cart results in an equal and opposite force back on her. In this case both forces act on the same system and, therefore, cancel. Thus internal forces (between components of a system) cancel. Choosing System 1 was crucial to solving this problem.

Example:**Force on the Cart—Choosing a New System**

Calculate the force the professor exerts on the cart in [\[link\]](#) using data from the previous example if needed.

Strategy

If we now define the system of interest to be the cart plus equipment (System 2 in [\[link\]](#)), then the net external force on System 2 is the force the professor exerts on the cart minus friction. The force she exerts on the cart, \mathbf{F}_{prof} , is an external force acting on System 2. \mathbf{F}_{prof} was internal to System 1, but it is external to System 2 and will enter Newton's second law for System 2.

Solution

Newton's second law can be used to find \mathbf{F}_{prof} . Starting with

Equation:

$$a = \frac{F_{\text{net}}}{m}$$

and noting that the magnitude of the net external force on System 2 is

Equation:

$$F_{\text{net}} = F_{\text{prof}} - f,$$

we solve for F_{prof} , the desired quantity:

Equation:

$$F_{\text{prof}} = F_{\text{net}} + f.$$

The value of f is given, so we must calculate net F_{net} . That can be done since both the acceleration and mass of System 2 are known. Using Newton's second law we see that

Equation:

$$F_{\text{net}} = ma,$$

where the mass of System 2 is 19.0 kg ($m = 12.0 \text{ kg} + 7.0 \text{ kg}$) and its acceleration was found to be $a = 1.5 \text{ m/s}^2$ in the previous example. Thus,

Equation:

$$F_{\text{net}} = ma,$$

Equation:

$$F_{\text{net}} = (19.0 \text{ kg})(1.5 \text{ m/s}^2) = 29 \text{ N}.$$

Now we can find the desired force:

Equation:

$$F_{\text{prof}} = F_{\text{net}} + f,$$

Equation:

$$F_{\text{prof}} = 29 \text{ N} + 24.0 \text{ N} = 53 \text{ N}.$$

Discussion

It is interesting that this force is significantly less than the 150-N force the professor exerted backward on the floor. Not all of that 150-N force is transmitted to the cart; some of it accelerates the professor.

The choice of a system is an important analytical step both in solving problems and in thoroughly understanding the physics of the situation (which is not necessarily the same thing).

Note:

PhET Explorations: Gravity Force Lab

Visualize the gravitational force that two objects exert on each other. Change properties of the objects in order to see how it changes the gravity force.

https://phet.colorado.edu/sims/html/gravity-force-lab/latest/gravity-force-lab_en.html

Section Summary

- **Newton's third law of motion** represents a basic symmetry in nature. It states: Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that the first body exerts.
- A **thrust** is a reaction force that pushes a body forward in response to a backward force. Rockets, airplanes, and cars are pushed forward by a thrust reaction force.

Conceptual Questions

Exercise:

Problem:

When you take off in a jet aircraft, there is a sensation of being pushed back into the seat. Explain why you move backward in the seat—is there really a force backward on you? (The same reasoning explains whiplash injuries, in which the head is apparently thrown backward.)

Exercise:

Problem:

A device used since the 1940s to measure the kick or recoil of the body due to heart beats is the “ballistocardiograph.” What physics principle(s) are involved here to measure the force of cardiac contraction? How might we construct such a device?

Exercise:

Problem:

Describe a situation in which one system exerts a force on another and, as a consequence, experiences a force that is equal in magnitude and opposite in direction. Which of Newton's laws of motion apply?

Exercise:**Problem:**

Why does an ordinary rifle recoil (kick backward) when fired? The barrel of a recoilless rifle is open at both ends. Describe how Newton's third law applies when one is fired. Can you safely stand close behind one when it is fired?

Exercise:**Problem:**

An American football lineman reasons that it is senseless to try to out-push the opposing player, since no matter how hard he pushes he will experience an equal and opposite force from the other player. Use Newton's laws and draw a free-body diagram of an appropriate system to explain how he can still out-push the opposition if he is strong enough.

Exercise:**Problem:**

Newton's third law of motion tells us that forces always occur in pairs of equal and opposite magnitude. Explain how the choice of the "system of interest" affects whether one such pair of forces cancels.

Problem Exercises**Exercise:**

Problem:

What net external force is exerted on a 1100-kg artillery shell fired from a battleship if the shell is accelerated at $2.40 \times 10^4 \text{ m/s}^2$? What is the magnitude of the force exerted on the ship by the artillery shell?

Solution:

Force on shell: $2.64 \times 10^7 \text{ N}$

Force exerted on ship = $-2.64 \times 10^7 \text{ N}$, by Newton's third law

Exercise:**Problem:**

A brave but inadequate rugby player is being pushed backward by an opposing player who is exerting a force of 800 N on him. The mass of the losing player plus equipment is 90.0 kg, and he is accelerating at 1.20 m/s^2 backward. (a) What is the force of friction between the losing player's feet and the grass? (b) What force does the winning player exert on the ground to move forward if his mass plus equipment is 110 kg? (c) Draw a sketch of the situation showing the system of interest used to solve each part. For this situation, draw a free-body diagram and write the net force equation.

Glossary**Newton's third law of motion**

whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that the first body exerts

thrust

a reaction force that pushes a body forward in response to a backward force; rockets, airplanes, and cars are pushed forward by a thrust reaction force

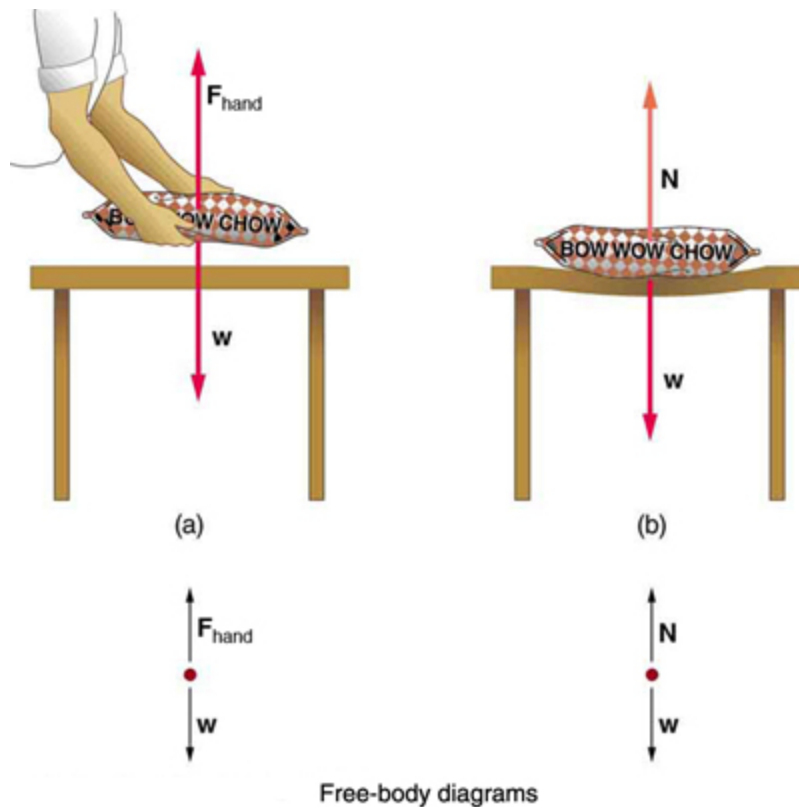
Normal, Tension, and Other Examples of Forces

- Define normal and tension forces.
- Apply Newton's laws of motion to solve problems involving a variety of forces.
- Use trigonometric identities to resolve weight into components.

Forces are given many names, such as push, pull, thrust, lift, weight, friction, and tension. Traditionally, forces have been grouped into several categories and given names relating to their source, how they are transmitted, or their effects. The most important of these categories are discussed in this section, together with some interesting applications. Further examples of forces are discussed later in this text.

Normal Force

Weight (also called force of gravity) is a pervasive force that acts at all times and must be counteracted to keep an object from falling. You definitely notice that you must support the weight of a heavy object by pushing up on it when you hold it stationary, as illustrated in [\[link\]](#)(a). But how do inanimate objects like a table support the weight of a mass placed on them, such as shown in [\[link\]](#)(b)? When the bag of dog food is placed on the table, the table actually sags slightly under the load. This would be noticeable if the load were placed on a card table, but even rigid objects deform when a force is applied to them. Unless the object is deformed beyond its limit, it will exert a restoring force much like a deformed spring (or trampoline or diving board). The greater the deformation, the greater the restoring force. So when the load is placed on the table, the table sags until the restoring force becomes as large as the weight of the load. At this point the net external force on the load is zero. That is the situation when the load is stationary on the table. The table sags quickly, and the sag is slight so we do not notice it. But it is similar to the sagging of a trampoline when you climb onto it.



(a) The person holding the bag of dog food must supply an upward force \mathbf{F}_{hand} equal in magnitude and opposite in direction to the weight of the food \mathbf{w} . (b) The card table sags when the dog food is placed on it, much like a stiff trampoline. Elastic restoring forces in the table grow as it sags until they supply a force \mathbf{N} equal in magnitude and opposite in direction to the weight of the load.

We must conclude that whatever supports a load, be it animate or not, must supply an upward force equal to the weight of the load, as we assumed in a few of the previous examples. If the force supporting a load is perpendicular to the surface of contact between the load and its support, this force is defined to be a **normal force** and here is given the symbol \mathbf{N} . (This is not the unit for force N.) The word *normal* means perpendicular to a

surface. The normal force can be less than the object's weight if the object is on an incline, as you will see in the next example.

Note:

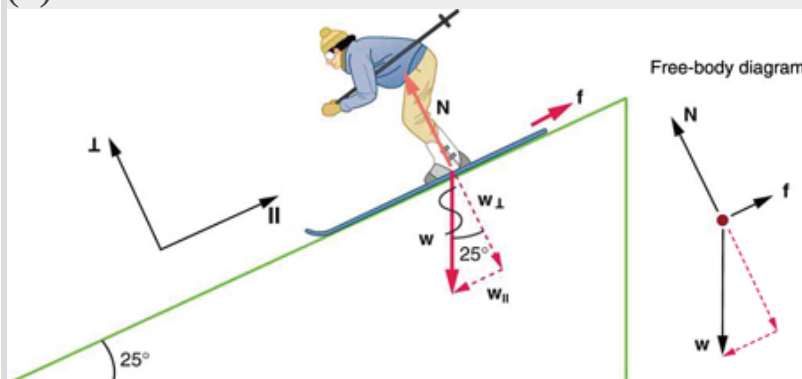
Common Misconception: Normal Force (\mathbf{N}) vs. Newton (N)

In this section we have introduced the quantity normal force, which is represented by the variable \mathbf{N} . This should not be confused with the symbol for the newton, which is also represented by the letter N . These symbols are particularly important to distinguish because the units of a normal force (\mathbf{N}) happen to be newtons (N). For example, the normal force \mathbf{N} that the floor exerts on a chair might be $\mathbf{N} = 100 \text{ N}$. One important difference is that normal force is a vector, while the newton is simply a unit. Be careful not to confuse these letters in your calculations! You will encounter more similarities among variables and units as you proceed in physics. Another example of this is the quantity work (W) and the unit watts (W).

Example:

Weight on an Incline, a Two-Dimensional Problem

Consider the skier on a slope shown in [\[link\]](#). Her mass including equipment is 60.0 kg . (a) What is her acceleration if friction is negligible? (b) What is her acceleration if friction is known to be 45.0 N ?



Since motion and friction are parallel to the slope, it is most convenient to project all

forces onto a coordinate system where one axis is parallel to the slope and the other is perpendicular (axes shown to left of skier).

\mathbf{N} is perpendicular to the slope and \mathbf{f} is parallel to the slope, but \mathbf{w} has components along both axes, namely \mathbf{w}_\perp and \mathbf{w}_\parallel . \mathbf{N} is equal in magnitude to \mathbf{w}_\perp , so that there is no motion perpendicular to the slope, but f is less than w_\parallel , so that there is a downslope acceleration (along the parallel axis).

Strategy

This is a two-dimensional problem, since the forces on the skier (the system of interest) are not parallel. The approach we have used in two-dimensional kinematics also works very well here. Choose a convenient coordinate system and project the vectors onto its axes, creating two connected *one*-dimensional problems to solve. The most convenient coordinate system for motion on an incline is one that has one coordinate parallel to the slope and one perpendicular to the slope. (Remember that motions along mutually perpendicular axes are independent.) We use the symbols \perp and \parallel to represent perpendicular and parallel, respectively. This choice of axes simplifies this type of problem, because there is no motion perpendicular to the slope and because friction is always parallel to the surface between two objects. The only external forces acting on the system are the skier's weight, friction, and the support of the slope, respectively labeled \mathbf{w} , \mathbf{f} , and \mathbf{N} in [\[link\]](#). \mathbf{N} is always perpendicular to the slope, and \mathbf{f} is parallel to it. But \mathbf{w} is not in the direction of either axis, and so the first step we take is to project it into components along the chosen axes, defining w_\parallel to be the component of weight parallel to the slope and w_\perp the component of weight perpendicular to the slope. Once this is done, we can consider the two separate problems of forces parallel to the slope and forces perpendicular to the slope.

Solution

The magnitude of the component of the weight parallel to the slope is $w_\parallel = w \sin(25^\circ) = mg \sin(25^\circ)$, and the magnitude of the component of

the weight perpendicular to the slope is

$$w_{\perp} = w \cos (25^{\circ}) = mg \cos (25^{\circ}).$$

(a) Neglecting friction. Since the acceleration is parallel to the slope, we need only consider forces parallel to the slope. (Forces perpendicular to the slope add to zero, since there is no acceleration in that direction.) The forces parallel to the slope are the amount of the skier's weight parallel to the slope w_{\parallel} and friction f . Using Newton's second law, with subscripts to denote quantities parallel to the slope,

Equation:

$$a_{\parallel} = \frac{F_{\text{net}\parallel}}{m}$$

where $F_{\text{net}\parallel} = w_{\parallel} = mg \sin (25^{\circ})$, assuming no friction for this part, so that

Equation:

$$a_{\parallel} = \frac{F_{\text{net}\parallel}}{m} = \frac{mg \sin (25^{\circ})}{m} = g \sin (25^{\circ})$$

Equation:

$$(9.80 \text{ m/s}^2)(0.4226) = 4.14 \text{ m/s}^2$$

is the acceleration.

(b) Including friction. We now have a given value for friction, and we know its direction is parallel to the slope and it opposes motion between surfaces in contact. So the net external force is now

Equation:

$$F_{\text{net}\parallel} = w_{\parallel} - f,$$

and substituting this into Newton's second law, $a_{\parallel} = \frac{F_{\text{net}\parallel}}{m}$, gives

Equation:

$$a_{\parallel} = \frac{F_{\text{net}\parallel}}{m} = \frac{w_{\parallel} - f}{m} = \frac{mg \sin (25^{\circ}) - f}{m}.$$

We substitute known values to obtain

Equation:

$$a_{\parallel} = \frac{(60.0 \text{ kg})(9.80 \text{ m/s}^2)(0.4226) - 45.0 \text{ N}}{60.0 \text{ kg}},$$

which yields

Equation:

$$a_{\parallel} = 3.39 \text{ m/s}^2,$$

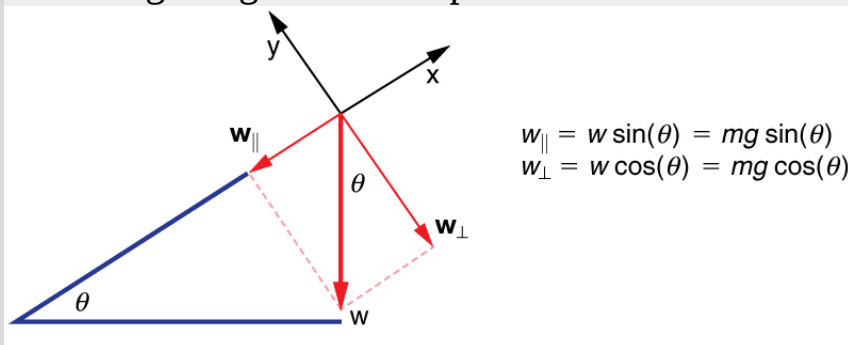
which is the acceleration parallel to the incline when there is 45.0 N of opposing friction.

Discussion

Since friction always opposes motion between surfaces, the acceleration is smaller when there is friction than when there is none. In fact, it is a general result that if friction on an incline is negligible, then the acceleration down the incline is $a = g \sin\theta$, *regardless of mass*. This is related to the previously discussed fact that all objects fall with the same acceleration in the absence of air resistance. Similarly, all objects, regardless of mass, slide down a frictionless incline with the same acceleration (if the angle is the same).

Note:

Resolving Weight into Components



An object rests on an incline that makes an

angle θ with the horizontal.

When an object rests on an incline that makes an angle θ with the horizontal, the force of gravity acting on the object is divided into two components: a force acting perpendicular to the plane, \mathbf{w}_{\perp} , and a force acting parallel to the plane, \mathbf{w}_{\parallel} . The perpendicular force of weight, \mathbf{w}_{\perp} , is typically equal in magnitude and opposite in direction to the normal force, \mathbf{N} . The force acting parallel to the plane, \mathbf{w}_{\parallel} , causes the object to accelerate down the incline. The force of friction, \mathbf{f} , opposes the motion of the object, so it acts upward along the plane.

It is important to be careful when resolving the weight of the object into components. If the angle of the incline is at an angle θ to the horizontal, then the magnitudes of the weight components are

Equation:

$$w_{\parallel} = w \sin (\theta) = mg \sin (\theta)$$

and

Equation:

$$w_{\perp} = w \cos (\theta) = mg \cos (\theta).$$

Instead of memorizing these equations, it is helpful to be able to determine them from reason. To do this, draw the right triangle formed by the three weight vectors. Notice that the angle θ of the incline is the same as the angle formed between \mathbf{w} and \mathbf{w}_{\perp} . Knowing this property, you can use trigonometry to determine the magnitude of the weight components:

Equation:

$$\begin{aligned} \cos (\theta) &= \frac{w_{\perp}}{w} \\ w_{\perp} &= w \cos (\theta) = mg \cos (\theta) \end{aligned}$$

Equation:

$$\begin{aligned} \sin (\theta) &= \frac{w_{\parallel}}{w} \\ w_{\parallel} &= w \sin (\theta) = mg \sin (\theta) \end{aligned}$$

Note:

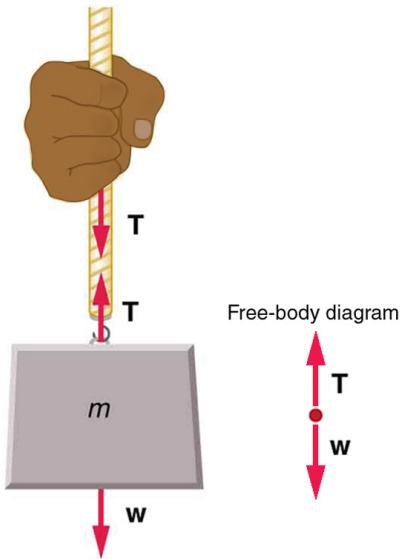
Take-Home Experiment: Force Parallel

To investigate how a force parallel to an inclined plane changes, find a rubber band, some objects to hang from the end of the rubber band, and a board you can position at different angles. How much does the rubber band stretch when you hang the object from the end of the board? Now place the board at an angle so that the object slides off when placed on the board. How much does the rubber band extend if it is lined up parallel to the board and used to hold the object stationary on the board? Try two more angles. What does this show?

Tension

A **tension** is a force along the length of a medium, especially a force carried by a flexible medium, such as a rope or cable. The word “tension” comes from a Latin word meaning “to stretch.” Not coincidentally, the flexible cords that carry muscle forces to other parts of the body are called *tendons*. Any flexible connector, such as a string, rope, chain, wire, or cable, can exert pulls only parallel to its length; thus, a force carried by a flexible connector is a tension with direction parallel to the connector. It is important to understand that tension is a pull in a connector. In contrast, consider the phrase: “You can’t push a rope.” The tension force pulls outward along the two ends of a rope.

Consider a person holding a mass on a rope as shown in [\[link\]](#).



When a perfectly flexible connector (one requiring no force to bend it) such as this rope transmits a force \mathbf{T} , that force must be parallel to the length of the rope, as shown. The pull such a flexible connector exerts is a tension. Note that the rope pulls with equal force but in opposite directions on the hand and the supported mass (neglecting the weight of the rope). This is an example of Newton's third law. The rope is the medium that carries

the equal and opposite forces between the two objects. The tension anywhere in the rope between the hand and the mass is equal. Once you have determined the tension in one location, you have determined the tension at all locations along the rope.

Tension in the rope must equal the weight of the supported mass, as we can prove using Newton's second law. If the 5.00-kg mass in the figure is stationary, then its acceleration is zero, and thus $\mathbf{F}_{\text{net}} = 0$. The only external forces acting on the mass are its weight \mathbf{w} and the tension \mathbf{T} supplied by the rope. Thus,

Equation:

$$F_{\text{net}} = T - w = 0,$$

where T and w are the magnitudes of the tension and weight and their signs indicate direction, with up being positive here. Thus, just as you would expect, the tension equals the weight of the supported mass:

Equation:

$$T = w = mg.$$

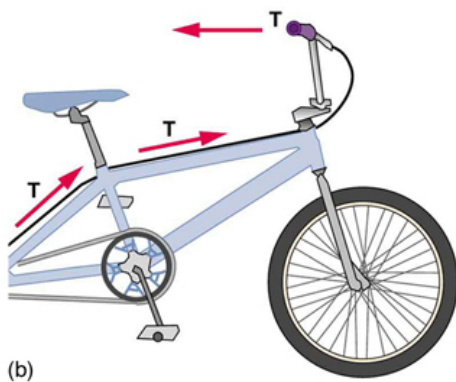
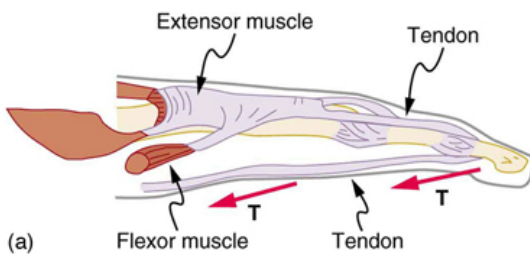
For a 5.00-kg mass, then (neglecting the mass of the rope) we see that

Equation:

$$T = mg = (5.00 \text{ kg})(9.80 \text{ m/s}^2) = 49.0 \text{ N}.$$

If we cut the rope and insert a spring, the spring would extend a length corresponding to a force of 49.0 N, providing a direct observation and measure of the tension force in the rope.

Flexible connectors are often used to transmit forces around corners, such as in a hospital traction system, a finger joint, or a bicycle brake cable. If there is no friction, the tension is transmitted undiminished. Only its direction changes, and it is always parallel to the flexible connector. This is illustrated in [\[link\]](#) (a) and (b).



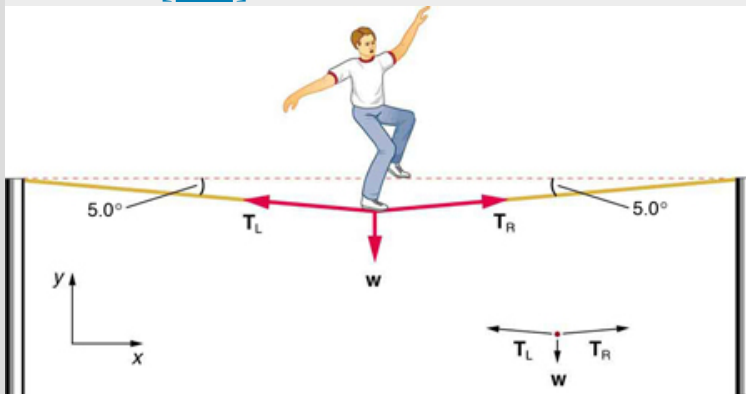
(a) Tendons in the finger carry force \mathbf{T} from the muscles to other parts of the finger, usually changing the force's direction, but not its magnitude (the

tendons are relatively friction free). (b) The brake cable on a bicycle carries the tension \mathbf{T} from the handlebars to the brake mechanism. Again, the direction but not the magnitude of \mathbf{T} is changed.

Example:

What Is the Tension in a Tightrope?

Calculate the tension in the wire supporting the 70.0-kg tightrope walker shown in [\[link\]](#).



The weight of a tightrope walker causes a wire to sag by 5.0 degrees. The system of interest here is the point in the wire at which the tightrope walker is standing.

Strategy

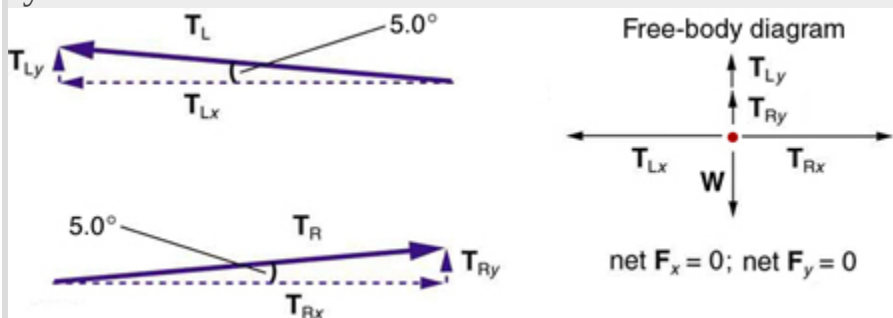
As you can see in the figure, the wire is not perfectly horizontal (it cannot be!), but is bent under the person's weight. Thus, the tension on either side of the person has an upward component that can support his weight. As

usual, forces are vectors represented pictorially by arrows having the same directions as the forces and lengths proportional to their magnitudes. The system is the tightrope walker, and the only external forces acting on him are his weight \mathbf{w} and the two tensions \mathbf{T}_L (left tension) and \mathbf{T}_R (right tension), as illustrated. It is reasonable to neglect the weight of the wire itself. The net external force is zero since the system is stationary. A little trigonometry can now be used to find the tensions. One conclusion is possible at the outset—we can see from part (b) of the figure that the magnitudes of the tensions T_L and T_R must be equal. This is because there is no horizontal acceleration in the rope, and the only forces acting to the left and right are T_L and T_R . Thus, the magnitude of those forces must be equal so that they cancel each other out.

Whenever we have two-dimensional vector problems in which no two vectors are parallel, the easiest method of solution is to pick a convenient coordinate system and project the vectors onto its axes. In this case the best coordinate system has one axis horizontal and the other vertical. We call the horizontal the x -axis and the vertical the y -axis.

Solution

First, we need to resolve the tension vectors into their horizontal and vertical components. It helps to draw a new free-body diagram showing all of the horizontal and vertical components of each force acting on the system.



When the vectors are projected onto vertical and horizontal axes, their components along those axes must add to zero, since the tightrope walker is stationary. The small angle results in T being much greater than w .

Consider the horizontal components of the forces (denoted with a subscript x):

Equation:

$$F_{\text{net}x} = T_{Lx} - T_{Rx}.$$

The net external horizontal force $F_{\text{net}x} = 0$, since the person is stationary. Thus,

Equation:

$$\begin{aligned} F_{\text{net}x} = 0 &= T_{Lx} - T_{Rx} \\ T_{Lx} &= T_{Rx}. \end{aligned}$$

Now, observe [\[link\]](#). You can use trigonometry to determine the magnitude of T_L and T_R . Notice that:

Equation:

$$\begin{aligned} \cos(5.0^\circ) &= \frac{T_{Lx}}{T_L} \\ T_{Lx} &= T_L \cos(5.0^\circ) \\ \cos(5.0^\circ) &= \frac{T_{Rx}}{T_R} \\ T_{Rx} &= T_R \cos(5.0^\circ). \end{aligned}$$

Equating T_{Lx} and T_{Rx} :

Equation:

$$T_L \cos(5.0^\circ) = T_R \cos(5.0^\circ).$$

Thus,

Equation:

$$T_L = T_R = T,$$

as predicted. Now, considering the vertical components (denoted by a subscript y), we can solve for T . Again, since the person is stationary, Newton's second law implies that net $F_y = 0$. Thus, as illustrated in the free-body diagram in [\[link\]](#),

Equation:

$$F_{\text{net}y} = T_{Ly} + T_{Ry} - w = 0.$$

Observing [\[link\]](#), we can use trigonometry to determine the relationship between T_{Ly} , T_{Ry} , and T . As we determined from the analysis in the horizontal direction, $T_L = T_R = T$:

Equation:

$$\begin{aligned}\sin (5.0^\circ) &= \frac{T_{Ly}}{T_L} \\ T_{Ly} = T_L \sin (5.0^\circ) &= T \sin (5.0^\circ) \\ \sin (5.0^\circ) &= \frac{T_{Ry}}{T_R} \\ T_{Ry} = T_R \sin (5.0^\circ) &= T \sin (5.0^\circ).\end{aligned}$$

Now, we can substitute the values for T_{Ly} and T_{Ry} , into the net force equation in the vertical direction:

Equation:

$$\begin{aligned}F_{\text{net}y} &= T_{Ly} + T_{Ry} - w = 0 \\ F_{\text{net}y} &= T \sin (5.0^\circ) + T \sin (5.0^\circ) - w = 0 \\ 2 T \sin (5.0^\circ) - w &= 0 \\ 2 T \sin (5.0^\circ) &= w\end{aligned}$$

and

Equation:

$$T = \frac{w}{2 \sin (5.0^\circ)} = \frac{mg}{2 \sin (5.0^\circ)},$$

so that

Equation:

$$T = \frac{(70.0 \text{ kg})(9.80 \text{ m/s}^2)}{2(0.0872)},$$

and the tension is

Equation:

$$T = 3900 \text{ N}.$$

Discussion

Note that the vertical tension in the wire acts as a normal force that supports the weight of the tightrope walker. The tension is almost six times the 686-N weight of the tightrope walker. Since the wire is nearly horizontal, the vertical component of its tension is only a small fraction of the tension in the wire. The large horizontal components are in opposite directions and cancel, and so most of the tension in the wire is not used to support the weight of the tightrope walker.

If we wish to *create* a very large tension, all we have to do is exert a force perpendicular to a flexible connector, as illustrated in [\[link\]](#). As we saw in the last example, the weight of the tightrope walker acted as a force perpendicular to the rope. We saw that the tension in the rope related to the weight of the tightrope walker in the following way:

Equation:

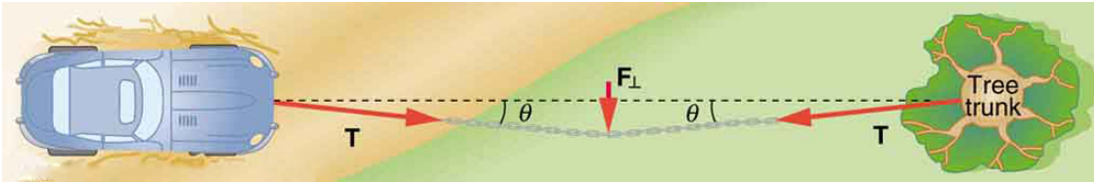
$$T = \frac{w}{2 \sin (\theta)}.$$

We can extend this expression to describe the tension T created when a perpendicular force (\mathbf{F}_\perp) is exerted at the middle of a flexible connector:

Equation:

$$T = \frac{F_\perp}{2 \sin (\theta)}.$$

Note that θ is the angle between the horizontal and the bent connector. In this case, T becomes very large as θ approaches zero. Even the relatively small weight of any flexible connector will cause it to sag, since an infinite tension would result if it were horizontal (i.e., $\theta = 0$ and $\sin \theta = 0$). (See [\[link\]](#).)



We can create a very large tension in the chain by pushing on it perpendicular to its length, as shown. Suppose we wish to pull a car out of the mud when no tow truck is available. Each time the car moves forward, the chain is tightened to keep it as nearly straight as possible. The tension in the chain is given by $T = \frac{F_{\perp}}{2 \sin(\theta)}$; since θ is small, T is very large. This situation is analogous to the tightrope walker shown in [\[link\]](#), except that the tensions shown here are those transmitted to the car and the tree rather than those acting at the point where \mathbf{F}_{\perp} is applied.



Unless an infinite tension is exerted, any flexible connector—such as the chain at the bottom of the picture—will sag under its own weight, giving a characteristic curve when the weight is evenly

distributed along the length.

Suspension bridges—such as the Golden Gate Bridge shown in this image—are essentially very heavy flexible connectors. The weight of the bridge is evenly distributed along the length of flexible connectors, usually cables, which take on the characteristic shape.

(credit: Leaflet, Wikimedia Commons)

Extended Topic: Real Forces and Inertial Frames

There is another distinction among forces in addition to the types already mentioned. Some forces are real, whereas others are not. *Real forces* are those that have some physical origin, such as the gravitational pull. Contrastingly, *fictitious forces* are those that arise simply because an observer is in an accelerating frame of reference, such as one that rotates (like a merry-go-round) or undergoes linear acceleration (like a car slowing down). For example, if a satellite is heading due north above Earth's northern hemisphere, then to an observer on Earth it will appear to experience a force to the west that has no physical origin. Of course, what is happening here is that Earth is rotating toward the east and moves east under the satellite. In Earth's frame this looks like a westward force on the satellite, or it can be interpreted as a violation of Newton's first law (the law of inertia). An **inertial frame of reference** is one in which all forces are real and, equivalently, one in which Newton's laws have the simple forms given in this chapter.

Earth's rotation is slow enough that Earth is nearly an inertial frame. You ordinarily must perform precise experiments to observe fictitious forces and the slight departures from Newton's laws, such as the effect just described. On the large scale, such as for the rotation of weather systems and ocean currents, the effects can be easily observed.

The crucial factor in determining whether a frame of reference is inertial is whether it accelerates or rotates relative to a known inertial frame. Unless stated otherwise, all phenomena discussed in this text are considered in inertial frames.

All the forces discussed in this section are real forces, but there are a number of other real forces, such as lift and thrust, that are not discussed in this section. They are more specialized, and it is not necessary to discuss every type of force. It is natural, however, to ask where the basic simplicity we seek to find in physics is in the long list of forces. Are some more basic than others? Are some different manifestations of the same underlying force? The answer to both questions is yes, as will be seen in the next (extended) section and in the treatment of modern physics later in the text.

Note:**PhET Explorations: Forces in 1 Dimension**

Explore the forces at work when you try to push a filing cabinet. Create an applied force and see the resulting friction force and total force acting on the cabinet. Charts show the forces, position, velocity, and acceleration vs. time. View a free-body diagram of all the forces (including gravitational and normal forces).

[Forces in
1
Dimension](#)

Section Summary

- When objects rest on a surface, the surface applies a force to the object that supports the weight of the object. This supporting force acts

perpendicular to and away from the surface. It is called a normal force, **N**.

- When objects rest on a non-accelerating horizontal surface, the magnitude of the normal force is equal to the weight of the object:

Equation:

$$N = mg.$$

- When objects rest on an inclined plane that makes an angle θ with the horizontal surface, the weight of the object can be resolved into components that act perpendicular (\mathbf{w}_\perp) and parallel (\mathbf{w}_\parallel) to the surface of the plane. These components can be calculated using:

Equation:

$$w_\parallel = w \sin(\theta) = mg \sin(\theta)$$

Equation:

$$w_\perp = w \cos(\theta) = mg \cos(\theta).$$

- The pulling force that acts along a stretched flexible connector, such as a rope or cable, is called tension, **T**. When a rope supports the weight of an object that is at rest, the tension in the rope is equal to the weight of the object:

Equation:

$$T = mg.$$

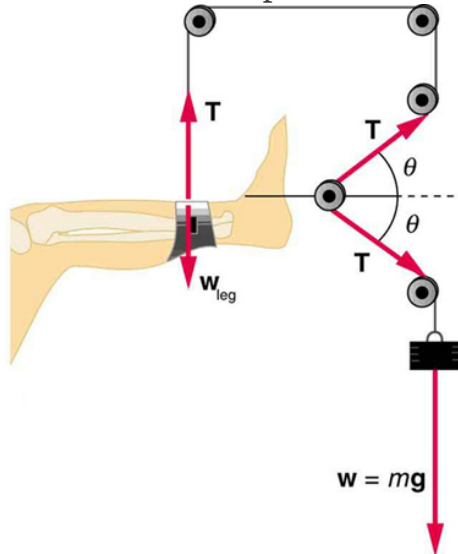
- In any inertial frame of reference (one that is not accelerated or rotated), Newton's laws have the simple forms given in this chapter and all forces are real forces having a physical origin.

Conceptual Questions

Exercise:

Problem:

If a leg is suspended by a traction setup as shown in [\[link\]](#), what is the tension in the rope?



A leg is suspended by a traction system in which wires are used to transmit forces. Frictionless pulleys change the direction of the force T without changing its magnitude.

Exercise:**Problem:**

In a traction setup for a broken bone, with pulleys and rope available, how might we be able to increase the force along the tibia using the same weight? (See [\[link\]](#).) (Note that the tibia is the shin bone shown in this image.)

Problem Exercises

Exercise:

Problem:

Two teams of nine members each engage in a tug of war. Each of the first team's members has an average mass of 68 kg and exerts an average force of 1350 N horizontally. Each of the second team's members has an average mass of 73 kg and exerts an average force of 1365 N horizontally. (a) What is magnitude of the acceleration of the two teams? (b) What is the tension in the section of rope between the teams?

Solution:

- a. 0.11 m/s^2
- b. $1.2 \times 10^4 \text{ N}$

Exercise:

Problem:

What force does a trampoline have to apply to a 45.0-kg gymnast to accelerate her straight up at 7.50 m/s^2 ? Note that the answer is independent of the velocity of the gymnast—she can be moving either up or down, or be stationary.

Exercise:

Problem:

(a) Calculate the tension in a vertical strand of spider web if a spider of mass $8.00 \times 10^{-5} \text{ kg}$ hangs motionless on it. (b) Calculate the tension in a horizontal strand of spider web if the same spider sits motionless in the middle of it much like the tightrope walker in [\[link\]](#). The strand sags at an angle of 12° below the horizontal. Compare this with the tension in the vertical strand (find their ratio).

Solution:

(a) $7.84 \times 10^{-4} \text{ N}$

(b) $1.89 \times 10^{-3} \text{ N}$. This is 2.41 times the tension in the vertical strand.

Exercise:**Problem:**

Suppose a 60.0-kg gymnast climbs a rope. (a) What is the tension in the rope if he climbs at a constant speed? (b) What is the tension in the rope if he accelerates upward at a rate of 1.50 m/s^2 ?

Exercise:**Problem:**

Show that, as stated in the text, a force \mathbf{F}_\perp exerted on a flexible medium at its center and perpendicular to its length (such as on the tightrope wire in [\[link\]](#)) gives rise to a tension of magnitude

$$T = \frac{F_\perp}{2 \sin(\theta)}.$$

Solution:

Newton's second law applied in vertical direction gives

Equation:

$$F_y = F - 2T \sin \theta = 0$$

Equation:

$$F = 2T \sin \theta$$

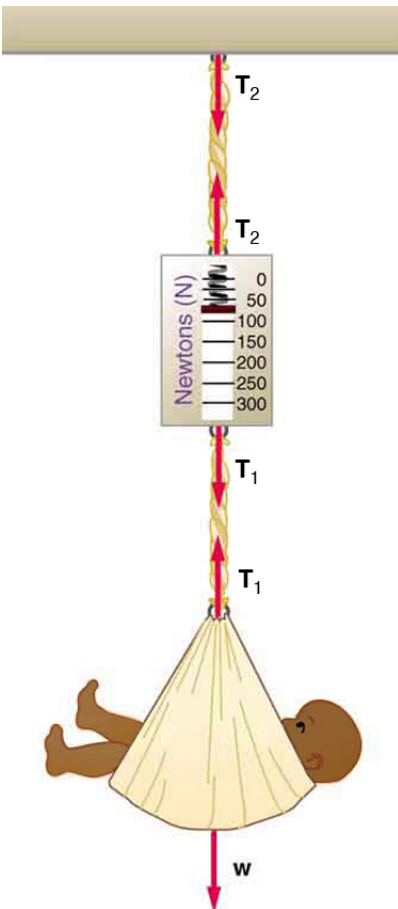
Equation:

$$T = \frac{F}{2 \sin \theta}.$$

Exercise:

Problem:

Consider the baby being weighed in [\[link\]](#). (a) What is the mass of the child and basket if a scale reading of 55 N is observed? (b) What is the tension T_1 in the cord attaching the baby to the scale? (c) What is the tension T_2 in the cord attaching the scale to the ceiling, if the scale has a mass of 0.500 kg? (d) Draw a sketch of the situation indicating the system of interest used to solve each part. The masses of the cords are negligible.



A baby is weighed
using a spring
scale.

Glossary

inertial frame of reference

a coordinate system that is not accelerating; all forces acting in an inertial frame of reference are real forces, as opposed to fictitious forces that are observed due to an accelerating frame of reference

normal force

the force that a surface applies to an object to support the weight of the object; acts perpendicular to the surface on which the object rests

tension

the pulling force that acts along a medium, especially a stretched flexible connector, such as a rope or cable; when a rope supports the weight of an object, the force on the object due to the rope is called a tension force

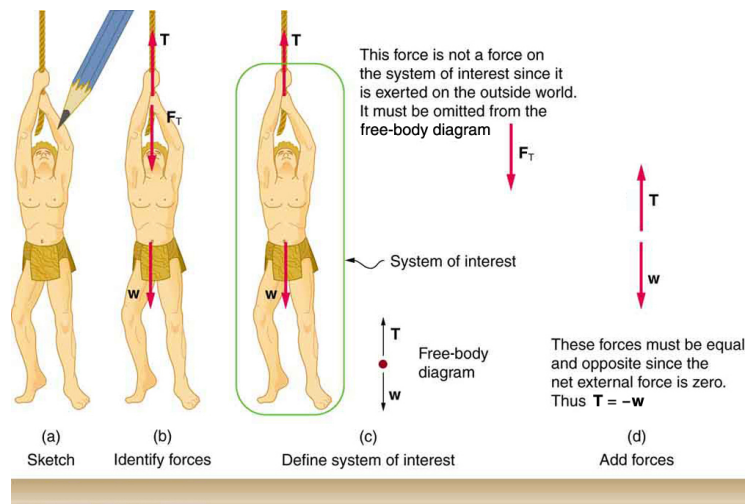
Problem-Solving Strategies

- Understand and apply a problem-solving procedure to solve problems using Newton's laws of motion.

Success in problem solving is obviously necessary to understand and apply physical principles, not to mention the more immediate need of passing exams. The basics of problem solving, presented earlier in this text, are followed here, but specific strategies useful in applying Newton's laws of motion are emphasized. These techniques also reinforce concepts that are useful in many other areas of physics. Many problem-solving strategies are stated outright in the worked examples, and so the following techniques should reinforce skills you have already begun to develop.

Problem-Solving Strategy for Newton's Laws of Motion

Step 1. As usual, it is first necessary to identify the physical principles involved. *Once it is determined that Newton's laws of motion are involved (if the problem involves forces), it is particularly important to draw a careful sketch of the situation.* Such a sketch is shown in [\[link\]\(a\)](#). Then, as in [\[link\]\(b\)](#), use arrows to represent all forces, label them carefully, and make their lengths and directions correspond to the forces they represent (whenever sufficient information exists).



(a) A sketch of Tarzan hanging from a vine. (b) Arrows are used to represent all forces. \mathbf{T} is the tension in the vine above Tarzan, \mathbf{F}_T is the force he exerts on the vine, and \mathbf{w} is his weight. All other forces, such as the nudge of a breeze, are assumed negligible. (c) Suppose we are given the ape man's mass and asked to find the tension in the vine. We then define the system of interest as shown and draw a free-body diagram. \mathbf{F}_T is no longer shown, because it is not a force acting on the system of interest; rather, \mathbf{F}_T acts on the outside world. (d) Showing only the arrows, the head-to-tail method of addition is used. It is apparent that $\mathbf{T} = -\mathbf{w}$, if Tarzan is stationary.

Step 2. Identify what needs to be determined and what is known or can be inferred from the problem as stated. That is, make a list of knowns and unknowns. *Then carefully determine the system of interest.* This decision is a crucial step, since Newton's second law involves only external forces. Once the system of interest has been identified, it becomes possible to determine which forces are external and which are internal, a necessary step to

employ Newton's second law. (See [\[link\]\(c\)](#).) Newton's third law may be used to identify whether forces are exerted between components of a system (internal) or between the system and something outside (external). As illustrated earlier in this chapter, the system of interest depends on what question we need to answer. This choice becomes easier with practice, eventually developing into an almost unconscious process. Skill in clearly defining systems will be beneficial in later chapters as well.

A diagram showing the system of interest and all of the external forces is called a **free-body diagram**. Only forces are shown on free-body diagrams, not acceleration or velocity. We have drawn several of these in worked examples. [\[link\]\(c\)](#) shows a free-body diagram for the system of interest. Note that no internal forces are shown in a free-body diagram.

Step 3. Once a free-body diagram is drawn, *Newton's second law can be applied to solve the problem*. This is done in [\[link\]\(d\)](#) for a particular situation. In general, once external forces are clearly identified in free-body diagrams, it should be a straightforward task to put them into equation form and solve for the unknown, as done in all previous examples. If the problem is one-dimensional—that is, if all forces are parallel—then they add like scalars. If the problem is two-dimensional, then it must be broken down into a pair of one-dimensional problems. This is done by projecting the force vectors onto a set of axes chosen for convenience. As seen in previous examples, the choice of axes can simplify the problem. For example, when an incline is involved, a set of axes with one axis parallel to the incline and one perpendicular to it is most convenient. It is almost always convenient to make one axis parallel to the direction of motion, if this is known.

Note:

Applying Newton's Second Law

Before you write net force equations, it is critical to determine whether the system is accelerating in a particular direction. If the acceleration is zero in a particular direction, then the net force is zero in that direction. Similarly, if the acceleration is nonzero in a particular direction, then the net force is described by the equation: $F_{\text{net}} = ma$. For example, if the system is accelerating in the horizontal direction, but it is not accelerating in the vertical direction, then you will have the following conclusions:

Equation:

$$F_{\text{net } x} = ma,$$

Equation:

$$F_{\text{net } y} = 0.$$

You will need this information in order to determine unknown forces acting in a system.

Step 4. As always, *check the solution to see whether it is reasonable*. In some cases, this is obvious. For example, it is reasonable to find that friction causes an object to slide down an incline more slowly than when no friction exists. In practice, intuition develops gradually through problem solving, and with experience it becomes progressively easier to judge whether an answer is reasonable. Another way to check your solution is to check the units. If you are solving for force and end up with units of m/s, then you have made a mistake.

Section Summary

- To solve problems involving Newton's laws of motion, follow the procedure described:
 1. Draw a sketch of the problem.
 2. Identify known and unknown quantities, and identify the system of interest. Draw a free-body diagram, which is a sketch showing all of the forces acting on an object. The object is represented by a dot, and the forces are represented by vectors extending in different directions from the dot. If vectors act in

directions that are not horizontal or vertical, resolve the vectors into horizontal and vertical components and draw them on the free-body diagram.

3. Write Newton's second law in the horizontal and vertical directions and add the forces acting on the object. If the object does not accelerate in a particular direction (for example, the x -direction) then $F_{\text{net } x} = 0$. If the object does accelerate in that direction, $F_{\text{net } x} = ma$.
4. Check your answer. Is the answer reasonable? Are the units correct?

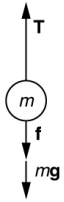
Problem Exercises

Exercise:

Problem:

A 5.00×10^5 -kg rocket is accelerating straight up. Its engines produce 1.250×10^7 N of thrust, and air resistance is 4.50×10^6 N. What is the rocket's acceleration? Explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion.

Solution:



Using the free-body diagram:

$$F_{\text{net}} = T - f - mg = ma,$$

so that

$$a = \frac{T - f - mg}{m} = \frac{1.250 \times 10^7 \text{ N} - 4.50 \times 10^6 \text{ N} - (5.00 \times 10^5 \text{ kg})(9.80 \text{ m/s}^2)}{5.00 \times 10^5 \text{ kg}} = 6.20 \text{ m/s}^2.$$

Exercise:

Problem:

The wheels of a midsize car exert a force of 2100 N backward on the road to accelerate the car in the forward direction. If the force of friction including air resistance is 250 N and the acceleration of the car is 1.80 m/s^2 , what is the mass of the car plus its occupants? Explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion. For this situation, draw a free-body diagram and write the net force equation.

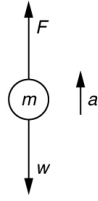
Exercise:

Problem:

Calculate the force a 70.0-kg high jumper must exert on the ground to produce an upward acceleration 4.00 times the acceleration due to gravity. Explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion.

Solution:

Use Newton's laws of motion.



Given : $a = 4.00g = (4.00)(9.80 \text{ m/s}^2) = 39.2 \text{ m/s}^2$; $m = 70.0 \text{ kg}$,

Find: F .

$$\sum F = +F - w = ma, \text{ so } F = ma + w = ma + mg = m(a + g).$$

that

$$F = (70.0 \text{ kg})[(39.2 \text{ m/s}^2) + (9.80 \text{ m/s}^2)] = 3.43 \times 10^3 \text{ N}.$$

The force exerted by the high-jumper is actually down on the ground, but F is up from the ground and makes him jump.

This result is reasonable, since it is quite possible for a person to exert a force of the magnitude of 10^3 N .

Exercise:

Problem:

When landing after a spectacular somersault, a 40.0-kg gymnast decelerates by pushing straight down on the mat. Calculate the force she must exert if her deceleration is 7.00 times the acceleration due to gravity.

Explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion.

Exercise:

Problem:

A freight train consists of two 8.00×10^4 -kg engines and 45 cars with average masses of $5.50 \times 10^4 \text{ kg}$. (a) What force must each engine exert backward on the track to accelerate the train at a rate of $5.00 \times 10^{-2} \text{ m/s}^2$ if the force of friction is $7.50 \times 10^5 \text{ N}$, assuming the engines exert identical forces? This is not a large frictional force for such a massive system. Rolling friction for trains is small, and consequently trains are very energy-efficient transportation systems. (b) What is the force in the coupling between the 37th and 38th cars (this is the force each exerts on the other), assuming all cars have the same mass and that friction is evenly distributed among all of the cars and engines?

Solution:

(a) $4.41 \times 10^5 \text{ N}$

(b) $1.50 \times 10^5 \text{ N}$

Exercise:

Problem:

Commercial airplanes are sometimes pushed out of the passenger loading area by a tractor. (a) An 1800-kg tractor exerts a force of $1.75 \times 10^4 \text{ N}$ backward on the pavement, and the system experiences forces resisting motion that total 2400 N. If the acceleration is 0.150 m/s^2 , what is the mass of the airplane? (b) Calculate the force exerted by the tractor on the airplane, assuming 2200 N of the friction is experienced by the airplane. (c) Draw two sketches showing the systems of interest used to solve each part, including the free-body diagrams for each.

Exercise:

Problem:

A 1100-kg car pulls a boat on a trailer. (a) What total force resists the motion of the car, boat, and trailer, if the car exerts a 1900-N force on the road and produces an acceleration of 0.550 m/s^2 ? The mass of the boat plus trailer is 700 kg. (b) What is the force in the hitch between the car and the trailer if 80% of the resisting forces are experienced by the boat and trailer?

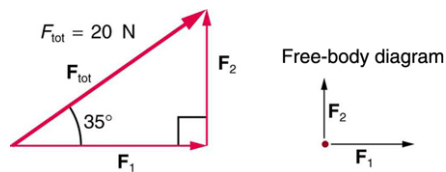
Solution:

(a) 910 N

(b) $1.11 \times 10^3 \text{ N}$

Exercise:**Problem:**

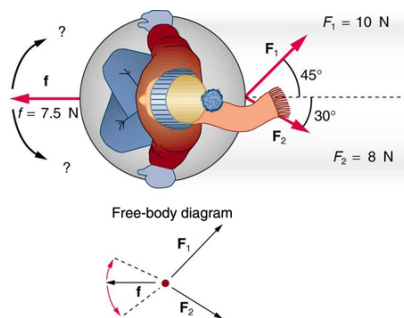
(a) Find the magnitudes of the forces \mathbf{F}_1 and \mathbf{F}_2 that add to give the total force \mathbf{F}_{tot} shown in [\[link\]](#). This may be done either graphically or by using trigonometry. (b) Show graphically that the same total force is obtained independent of the order of addition of \mathbf{F}_1 and \mathbf{F}_2 . (c) Find the direction and magnitude of some other pair of vectors that add to give \mathbf{F}_{tot} . Draw these to scale on the same drawing used in part (b) or a similar picture.

**Exercise:****Problem:**

Two children pull a third child on a snow saucer sled exerting forces \mathbf{F}_1 and \mathbf{F}_2 as shown from above in [\[link\]](#). Find the acceleration of the 49.00-kg sled and child system. Note that the direction of the frictional force is unspecified; it will be in the opposite direction of the sum of \mathbf{F}_1 and \mathbf{F}_2 .

Solution:

$a = 0.139 \text{ m/s}$, $\theta = 12.4^\circ$ north of east



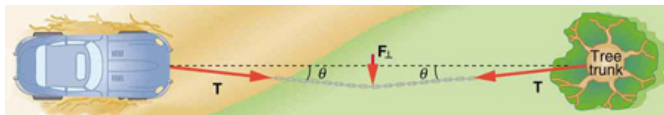
An overhead view of the horizontal forces acting on a

child's snow saucer sled.

Exercise:

Problem:

Suppose your car was mired deeply in the mud and you wanted to use the method illustrated in [\[link\]](#) to pull it out. (a) What force would you have to exert perpendicular to the center of the rope to produce a force of 12,000 N on the car if the angle is 2.00° ? In this part, explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion. (b) Real ropes stretch under such forces. What force would be exerted on the car if the angle increases to 7.00° and you still apply the force found in part (a) to its center?



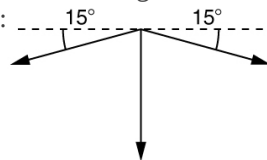
Exercise:

Problem:

What force is exerted on the tooth in [\[link\]](#) if the tension in the wire is 25.0 N? Note that the force applied to the tooth is smaller than the tension in the wire, but this is necessitated by practical considerations of how force can be applied in the mouth. Explicitly show how you follow steps in the Problem-Solving Strategy for Newton's laws of motion.

Solution:

Use Newton's laws since we are looking for forces.
Draw a free-body diagram:

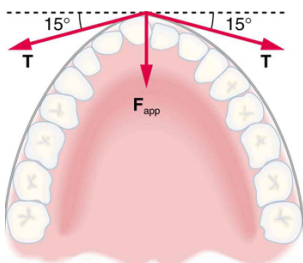


The tension is given as $T = 25.0 \text{ N}$. Find F_{app} . Using Newton's laws gives:

$\Sigma F_y = 0$, so that y-components of the two tensions is due to the applied force:

$$F_{\text{app}} = 2 T \sin \theta = 2(25.0 \text{ N}) \sin(15^\circ) =$$

This seems reasonable, since the applied tensions should be greater than the force applied to the tooth.



Braces are used to apply forces to teeth to realign them. Shown in this figure are the tensions applied by the wire to the protruding tooth. The total force applied to the tooth by the wire, \mathbf{F}_{app} , points straight toward the back of the mouth.

Exercise:

Problem:

[\[link\]](#) shows Superhero and Trusty Sidekick hanging motionless from a rope. Superhero's mass is 90.0 kg, while Trusty Sidekick's is 55.0 kg, and the mass of the rope is negligible. (a) Draw a free-body diagram of the situation showing all forces acting on Superhero, Trusty Sidekick, and the rope. (b) Find the tension in the rope above Superhero. (c) Find the tension in the rope between Superhero and Trusty Sidekick. Indicate on your free-body diagram the system of interest used to solve each part.



Superhero and Trusty Sidekick hang motionless on a rope as they try to figure out what to do next. Will the tension be the same everywhere in the rope?

Exercise:**Problem:**

A nurse pushes a cart by exerting a force on the handle at a downward angle 35.0° below the horizontal. The loaded cart has a mass of 28.0 kg, and the force of friction is 60.0 N. (a) Draw a free-body diagram for the system of interest. (b) What force must the nurse exert to move at a constant velocity?

Exercise:**Problem:**

Construct Your Own Problem Consider the tension in an elevator cable during the time the elevator starts from rest and accelerates its load upward to some cruising velocity. Taking the elevator and its load to be the system of interest, draw a free-body diagram. Then calculate the tension in the cable. Among the things to consider are the mass of the elevator and its load, the final velocity, and the time taken to reach that velocity.

Exercise:**Problem:**

Construct Your Own Problem Consider two people pushing a toboggan with four children on it up a snow-covered slope. Construct a problem in which you calculate the acceleration of the toboggan and its load. Include a free-body diagram of the appropriate system of interest as the basis for your analysis. Show vector forces and their components and explain the choice of coordinates. Among the things to be considered are the forces exerted by those pushing, the angle of the slope, and the masses of the toboggan and children.

Exercise:**Problem:**

Unreasonable Results (a) Repeat [\[link\]](#), but assume an acceleration of 1.20 m/s^2 is produced. (b) What is unreasonable about the result? (c) Which premise is unreasonable, and why is it unreasonable?

Exercise:**Problem:**

Unreasonable Results (a) What is the initial acceleration of a rocket that has a mass of $1.50 \times 10^6 \text{ kg}$ at takeoff, the engines of which produce a thrust of $2.00 \times 10^6 \text{ N}$? Do not neglect gravity. (b) What is unreasonable about the result? (This result has been unintentionally achieved by several real rockets.) (c) Which premise is unreasonable, or which premises are inconsistent? (You may find it useful to compare this problem to the rocket problem earlier in this section.)

Further Applications of Newton's Laws of Motion

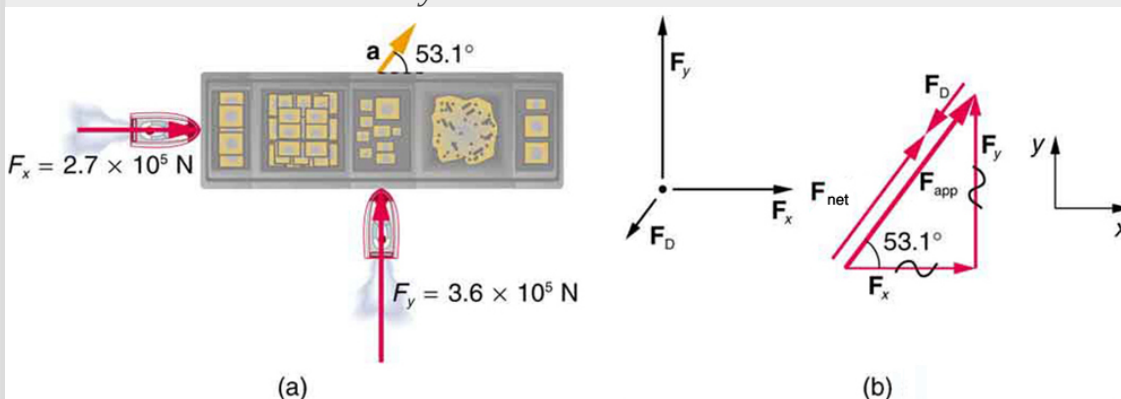
- Apply problem-solving techniques to solve for quantities in more complex systems of forces.
- Integrate concepts from kinematics to solve problems using Newton's laws of motion.

There are many interesting applications of Newton's laws of motion, a few more of which are presented in this section. These serve also to illustrate some further subtleties of physics and to help build problem-solving skills.

Example:

Drag Force on a Barge

Suppose two tugboats push on a barge at different angles, as shown in [\[link\]](#). The first tugboat exerts a force of $2.7 \times 10^5 \text{ N}$ in the x -direction, and the second tugboat exerts a force of $3.6 \times 10^5 \text{ N}$ in the y -direction.



(a) A view from above of two tugboats pushing on a barge. (b) The free-body diagram for the ship contains only forces acting in the plane of the water. It omits the two vertical forces—the weight of the barge and the buoyant force of the water supporting it cancel and are not shown. Since the applied forces are perpendicular, the x - and y -axes are in the same direction as \mathbf{F}_x and \mathbf{F}_y . The problem quickly becomes a one-dimensional problem along the direction of \mathbf{F}_{app} , since friction is in the direction opposite to \mathbf{F}_{app} .

If the mass of the barge is $5.0 \times 10^6 \text{ kg}$ and its acceleration is observed to be $7.5 \times 10^{-2} \text{ m/s}^2$ in the direction shown, what is the drag force of the water on the

barge resisting the motion? (Note: drag force is a frictional force exerted by fluids, such as air or water. The drag force opposes the motion of the object.)

Strategy

The directions and magnitudes of acceleration and the applied forces are given in [\[link\]\(a\)](#). We will define the total force of the tugboats on the barge as \mathbf{F}_{app} so that:

Equation:

$$\mathbf{F}_{\text{app}} = \mathbf{F}_x + \mathbf{F}_y$$

Since the barge is flat bottomed, the drag of the water \mathbf{F}_D will be in the direction opposite to \mathbf{F}_{app} , as shown in the free-body diagram in [\[link\]\(b\)](#). The system of interest here is the barge, since the forces on *it* are given as well as its acceleration. Our strategy is to find the magnitude and direction of the net applied force \mathbf{F}_{app} , and then apply Newton's second law to solve for the drag force \mathbf{F}_D .

Solution

Since \mathbf{F}_x and \mathbf{F}_y are perpendicular, the magnitude and direction of \mathbf{F}_{app} are easily found. First, the resultant magnitude is given by the Pythagorean theorem:

Equation:

$$\begin{aligned} F_{\text{app}} &= \sqrt{F_x^2 + F_y^2} \\ F_{\text{app}} &= \sqrt{(2.7 \times 10^5 \text{ N})^2 + (3.6 \times 10^5 \text{ N})^2} = 4.5 \times 10^5 \text{ N}. \end{aligned}$$

The angle is given by

Equation:

$$\begin{aligned} \theta &= \tan^{-1} \left(\frac{F_y}{F_x} \right) \\ \theta &= \tan^{-1} \left(\frac{3.6 \times 10^5 \text{ N}}{2.7 \times 10^5 \text{ N}} \right) = 53^\circ, \end{aligned}$$

which we know, because of Newton's first law, is the same direction as the acceleration. \mathbf{F}_D is in the opposite direction of \mathbf{F}_{app} , since it acts to slow down the acceleration. Therefore, the net external force is in the same direction as \mathbf{F}_{app} , but its magnitude is slightly less than \mathbf{F}_{app} . The problem is now one-dimensional. From [\[link\]\(b\)](#), we can see that

Equation:

$$F_{\text{net}} = F_{\text{app}} - F_D.$$

But Newton's second law states that

Equation:

$$F_{\text{net}} = ma.$$

Thus,

Equation:

$$F_{\text{app}} - F_{\text{D}} = ma.$$

This can be solved for the magnitude of the drag force of the water F_{D} in terms of known quantities:

Equation:

$$F_{\text{D}} = F_{\text{app}} - ma.$$

Substituting known values gives

Equation:

$$F_{\text{D}} = (4.5 \times 10^5 \text{ N}) - (5.0 \times 10^6 \text{ kg})(7.5 \times 10^{-2} \text{ m/s}^2) = 7.5 \times 10^4 \text{ N}.$$

The direction of \mathbf{F}_{D} has already been determined to be in the direction opposite to \mathbf{F}_{app} , or at an angle of 53° south of west.

Discussion

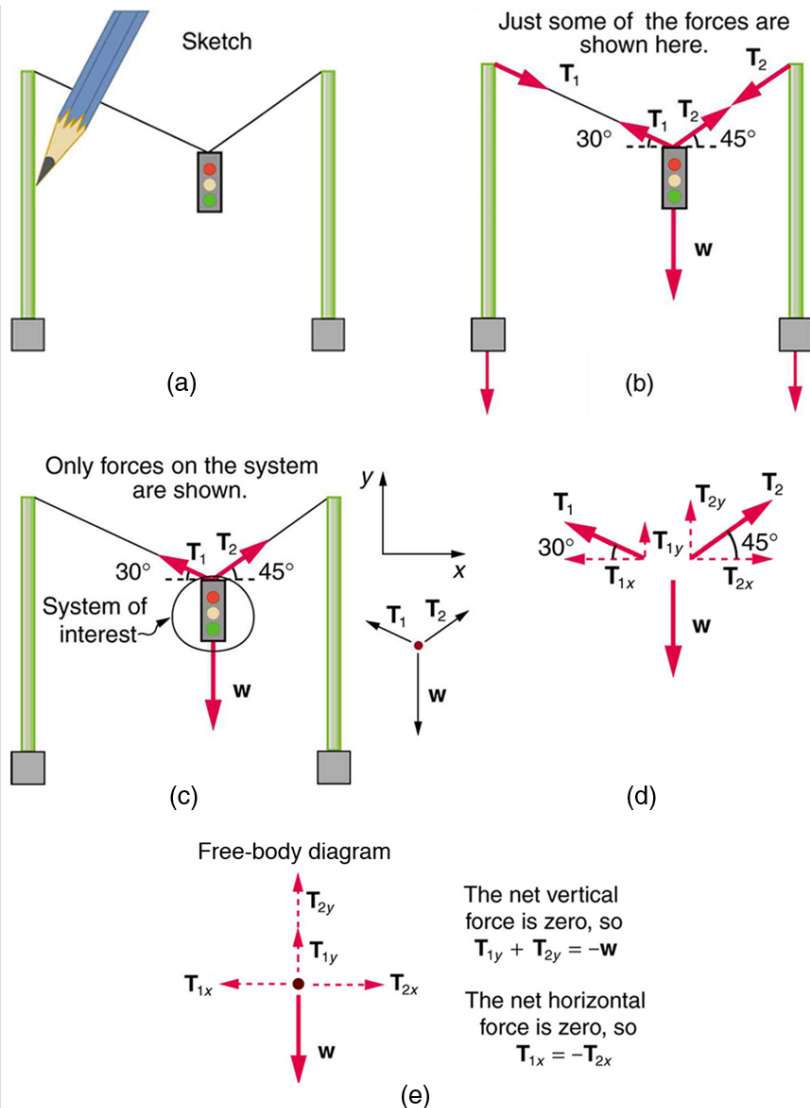
The numbers used in this example are reasonable for a moderately large barge. It is certainly difficult to obtain larger accelerations with tugboats, and small speeds are desirable to avoid running the barge into the docks. Drag is relatively small for a well-designed hull at low speeds, consistent with the answer to this example, where F_{D} is less than 1/600th of the weight of the ship.

In the earlier example of a tightrope walker we noted that the tensions in wires supporting a mass were equal only because the angles on either side were equal. Consider the following example, where the angles are not equal; slightly more trigonometry is involved.

Example:

Different Tensions at Different Angles

Consider the traffic light (mass 15.0 kg) suspended from two wires as shown in [\[link\]](#). Find the tension in each wire, neglecting the masses of the wires.



A traffic light is suspended from two wires. (b) Some of the forces involved. (c) Only forces acting on the system are shown here. The free-body diagram for the traffic light is also shown. (d) The forces projected onto vertical (y) and horizontal (x) axes. The horizontal components of the tensions must cancel, and the sum of the vertical components of the tensions must equal the weight of the traffic light. (e) The free-body diagram shows the vertical and horizontal forces acting on the traffic light.

Strategy

The system of interest is the traffic light, and its free-body diagram is shown in [\[link\]](#) (c). The three forces involved are not parallel, and so they must be projected onto a coordinate system. The most convenient coordinate system has one axis vertical and one horizontal, and the vector projections on it are shown in part (d) of the figure. There are two unknowns in this problem (T_1 and T_2), so two equations are needed to find them. These two equations come from applying Newton's second law along the vertical and horizontal axes, noting that the net external force is zero along each axis because acceleration is zero.

Solution

First consider the horizontal or x -axis:

Equation:

$$F_{\text{net}x} = T_{2x} - T_{1x} = 0.$$

Thus, as you might expect,

Equation:

$$T_{1x} = T_{2x}.$$

This gives us the following relationship between T_1 and T_2 :

Equation:

$$T_1 \cos(30^\circ) = T_2 \cos(45^\circ).$$

Thus,

Equation:

$$T_2 = (1.225)T_1.$$

Note that T_1 and T_2 are not equal in this case, because the angles on either side are not equal. It is reasonable that T_2 ends up being greater than T_1 , because it is exerted more vertically than T_1 .

Now consider the force components along the vertical or y -axis:

Equation:

$$F_{\text{net}y} = T_{1y} + T_{2y} - w = 0.$$

This implies

Equation:

$$T_{1y} + T_{2y} = w.$$

Substituting the expressions for the vertical components gives

Equation:

$$T_1 \sin (30^\circ) + T_2 \sin (45^\circ) = w.$$

There are two unknowns in this equation, but substituting the expression for T_2 in terms of T_1 reduces this to one equation with one unknown:

Equation:

$$T_1(0.500) + (1.225T_1)(0.707) = w = mg,$$

which yields

Equation:

$$(1.366)T_1 = (15.0 \text{ kg})(9.80 \text{ m/s}^2).$$

Solving this last equation gives the magnitude of T_1 to be

Equation:

$$T_1 = 108 \text{ N}.$$

Finally, the magnitude of T_2 is determined using the relationship between them, $T_2 = 1.225 T_1$, found above. Thus we obtain

Equation:

$$T_2 = 132 \text{ N}.$$

Discussion

Both tensions would be larger if both wires were more horizontal, and they will be equal if and only if the angles on either side are the same (as they were in the earlier example of a tightrope walker).

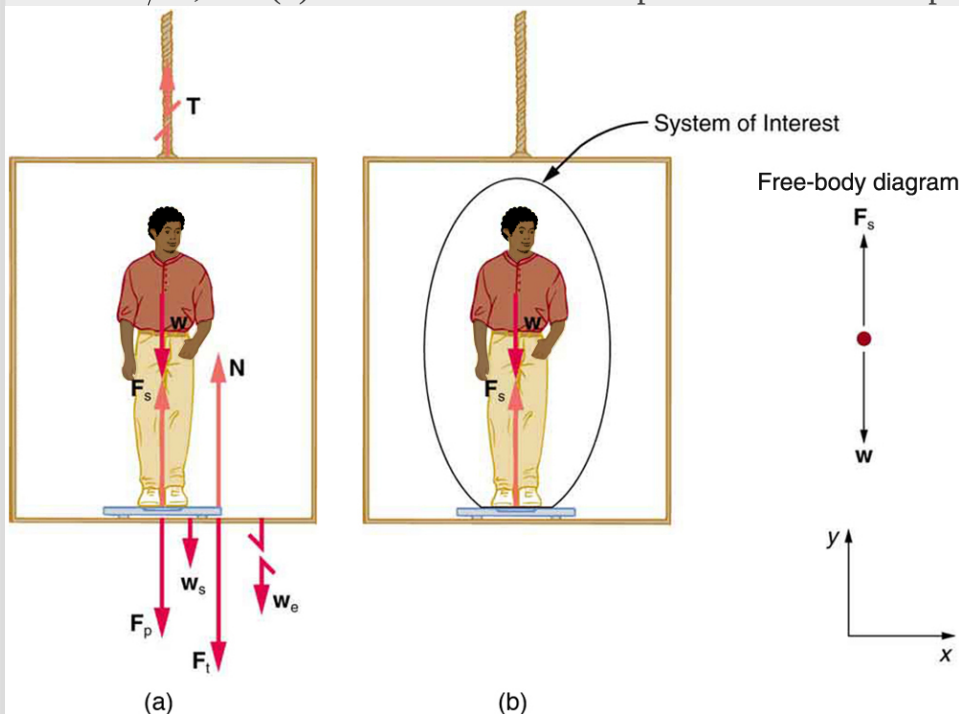
The bathroom scale is an excellent example of a normal force acting on a body. It provides a quantitative reading of how much it must push upward to support the weight of an object. But can you predict what you would see on the dial of a bathroom scale if you stood on it during an elevator ride? Will you see a value greater than your weight when the elevator starts up? What about when the elevator moves upward at a constant speed: will the scale still read more than your weight at rest? Consider the following example.

Example:

What Does the Bathroom Scale Read in an Elevator?

[\[link\]](#) shows a 75.0-kg man (weight of about 165 lb) standing on a bathroom scale in an elevator. Calculate the scale reading: (a) if the elevator accelerates upward at a rate

of 1.20 m/s^2 , and (b) if the elevator moves upward at a constant speed of 1 m/s .



(a) The various forces acting when a person stands on a bathroom scale in an elevator. The arrows are approximately correct for when the elevator is accelerating upward—broken arrows represent forces too large to be drawn to scale. \mathbf{T} is the tension in the supporting cable, \mathbf{w} is the weight of the person, \mathbf{w}_s is the weight of the scale, \mathbf{w}_e is the weight of the elevator, \mathbf{F}_s is the force of the scale on the person, \mathbf{F}_p is the force of the person on the scale, \mathbf{F}_t is the force of the scale on the floor of the elevator, and \mathbf{N} is the force of the floor upward on the scale. (b) The free-body diagram shows only the external forces acting on the designated system of interest—the person.

Strategy

If the scale is accurate, its reading will equal F_p , the magnitude of the force the person exerts downward on it. [\[link\]](#)(a) shows the numerous forces acting on the elevator, scale, and person. It makes this one-dimensional problem look much more formidable than if the person is chosen to be the system of interest and a free-body diagram is drawn as in [\[link\]](#)(b). Analysis of the free-body diagram using Newton's laws can produce answers to both parts (a) and (b) of this example, as well as some other questions that might arise. The only forces acting on the person are his weight \mathbf{w} and the upward force of the scale \mathbf{F}_s . According to Newton's third law \mathbf{F}_p and \mathbf{F}_s are

equal in magnitude and opposite in direction, so that we need to find F_s in order to find what the scale reads. We can do this, as usual, by applying Newton's second law,
Equation:

$$F_{\text{net}} = ma.$$

From the free-body diagram we see that $F_{\text{net}} = F_s - w$, so that
Equation:

$$F_s - w = ma.$$

Solving for F_s gives an equation with only one unknown:
Equation:

$$F_s = ma + w,$$

or, because $w = mg$, simply
Equation:

$$F_s = ma + mg.$$

No assumptions were made about the acceleration, and so this solution should be valid for a variety of accelerations in addition to the ones in this exercise.

Solution for (a)

In this part of the problem, $a = 1.20 \text{ m/s}^2$, so that

Equation:

$$F_s = (75.0 \text{ kg})(1.20 \text{ m/s}^2) + (75.0 \text{ kg})(9.80 \text{ m/s}^2),$$

yielding

Equation:

$$F_s = 825 \text{ N}.$$

Discussion for (a)

This is about 185 lb. What would the scale have read if he were stationary? Since his acceleration would be zero, the force of the scale would be equal to his weight:

Equation:

$$\begin{aligned} F_{\text{net}} &= ma = 0 = F_s - w \\ F_s &= w = mg \\ F_s &= (75.0 \text{ kg})(9.80 \text{ m/s}^2) \\ F_s &= 735 \text{ N}. \end{aligned}$$

So, the scale reading in the elevator is greater than his 735-N (165 lb) weight. This means that the scale is pushing up on the person with a force greater than his weight, as it must in order to accelerate him upward. Clearly, the greater the acceleration of the elevator, the greater the scale reading, consistent with what you feel in rapidly accelerating versus slowly accelerating elevators.

Solution for (b)

Now, what happens when the elevator reaches a constant upward velocity? Will the scale still read more than his weight? For any constant velocity—up, down, or stationary—acceleration is zero because $a = \frac{\Delta v}{\Delta t}$, and $\Delta v = 0$.

Thus,

Equation:

$$F_s = ma + mg = 0 + mg.$$

Now

Equation:

$$F_s = (75.0 \text{ kg})(9.80 \text{ m/s}^2),$$

which gives

Equation:

$$F_s = 735 \text{ N}.$$

Discussion for (b)

The scale reading is 735 N, which equals the person's weight. This will be the case whenever the elevator has a constant velocity—moving up, moving down, or stationary.

The solution to the previous example also applies to an elevator accelerating downward, as mentioned. When an elevator accelerates downward, a is negative, and the scale reading is *less* than the weight of the person, until a constant downward velocity is reached, at which time the scale reading again becomes equal to the person's weight. If the elevator is in free-fall and accelerating downward at g , then the scale reading will be zero and the person will *appear* to be weightless.

Integrating Concepts: Newton's Laws of Motion and Kinematics

Physics is most interesting and most powerful when applied to general situations that involve more than a narrow set of physical principles. Newton's laws of motion can also be integrated with other concepts that have been discussed previously in this text to

solve problems of motion. For example, forces produce accelerations, a topic of kinematics, and hence the relevance of earlier chapters. When approaching problems that involve various types of forces, acceleration, velocity, and/or position, use the following steps to approach the problem:

Problem-Solving Strategy

- Step 1. *Identify which physical principles are involved.* Listing the givens and the quantities to be calculated will allow you to identify the principles involved.
- Step 2. *Solve the problem using strategies outlined in the text.* If these are available for the specific topic, you should refer to them. You should also refer to the sections of the text that deal with a particular topic. The following worked example illustrates how these strategies are applied to an integrated concept problem.

Example:

What Force Must a Soccer Player Exert to Reach Top Speed?

A soccer player starts from rest and accelerates forward, reaching a velocity of 8.00 m/s in 2.50 s. (a) What was his average acceleration? (b) What average force did he exert backward on the ground to achieve this acceleration? The player’s mass is 70.0 kg, and air resistance is negligible.

Strategy

To solve an integrated concept problem, first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example considers acceleration along a straight line. This is a topic of kinematics. Part (b) deals with force, a dynamics topic found in this chapter.

The following solutions to each part of the example illustrate how the specific

problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so forth.

Solution for (a)

We are given the initial and final velocities (zero and 8.00 m/s forward); thus, the change in velocity is $\Delta v = 8.00 \text{ m/s}$. We are given the elapsed time, and so $\Delta t = 2.50 \text{ s}$. The unknown is acceleration, which can be found from its definition:

Equation:

$$a = \frac{\Delta v}{\Delta t}.$$

Substituting the known values yields

Equation:

$$\begin{aligned} a &= \frac{8.00 \text{ m/s}}{2.50 \text{ s}} \\ &= 3.20 \text{ m/s}^2. \end{aligned}$$

Discussion for (a)

This is an attainable acceleration for an athlete in good condition.

Solution for (b)

Here we are asked to find the average force the player exerts backward to achieve this forward acceleration. Neglecting air resistance, this would be equal in magnitude to the net external force on the player, since this force causes his acceleration. Since we now know the player's acceleration and are given his mass, we can use Newton's second law to find the force exerted. That is,

Equation:

$$F_{\text{net}} = ma.$$

Substituting the known values of m and a gives

Equation:

$$\begin{aligned} F_{\text{net}} &= (70.0 \text{ kg})(3.20 \text{ m/s}^2) \\ &= 224 \text{ N}. \end{aligned}$$

Discussion for (b)

This is about 50 pounds, a reasonable average force.

This worked example illustrates how to apply problem-solving strategies to situations that include topics from different chapters. The first step is to identify the physical principles involved in the problem. The second step is to solve for the unknown using familiar problem-solving strategies. These strategies are found throughout the text, and many worked examples show how to use them for single topics. You will find these

techniques for integrated concept problems useful in applications of physics outside of a physics course, such as in your profession, in other science disciplines, and in everyday life. The following problems will build your skills in the broad application of physical principles.

Summary

- Newton's laws of motion can be applied in numerous situations to solve problems of motion.
- Some problems will contain multiple force vectors acting in different directions on an object. Be sure to draw diagrams, resolve all force vectors into horizontal and vertical components, and draw a free-body diagram. Always analyze the direction in which an object accelerates so that you can determine whether $F_{\text{net}} = ma$ or $F_{\text{net}} = 0$.
- The normal force on an object is not always equal in magnitude to the weight of the object. If an object is accelerating, the normal force will be less than or greater than the weight of the object. Also, if the object is on an inclined plane, the normal force will always be less than the full weight of the object.
- Some problems will contain various physical quantities, such as forces, acceleration, velocity, or position. You can apply concepts from kinematics and dynamics in order to solve these problems of motion.

Conceptual Questions

Exercise:

Problem:

To simulate the apparent weightlessness of space orbit, astronauts are trained in the hold of a cargo aircraft that is accelerating downward at g . Why will they appear to be weightless, as measured by standing on a bathroom scale, in this accelerated frame of reference? Is there any difference between their apparent weightlessness in orbit and in the aircraft?

Exercise:

Problem:

A cartoon shows the toupee coming off the head of an elevator passenger when the elevator rapidly stops during an upward ride. Can this really happen without the person being tied to the floor of the elevator? Explain your answer.

Problem Exercises

Exercise:

Problem:

A flea jumps by exerting a force of 1.20×10^{-5} N straight down on the ground. A breeze blowing on the flea parallel to the ground exerts a force of 0.500×10^{-6} N on the flea. Find the direction and magnitude of the acceleration of the flea if its mass is 6.00×10^{-7} kg. Do not neglect the gravitational force.

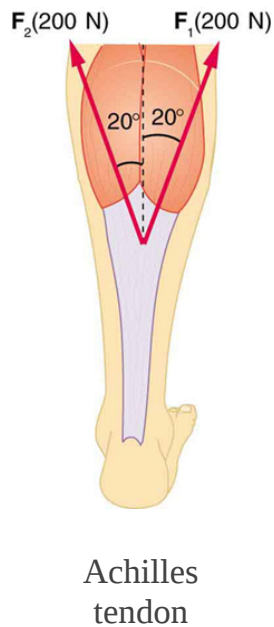
Solution:

10.2 m/s^2 , 4.67° from vertical

Exercise:

Problem:

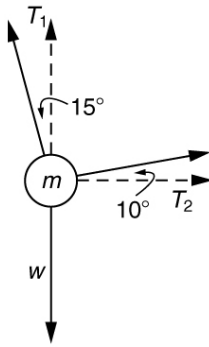
Two muscles in the back of the leg pull upward on the Achilles tendon, as shown in [\[link\]](#). (These muscles are called the medial and lateral heads of the gastrocnemius muscle.) Find the magnitude and direction of the total force on the Achilles tendon. What type of movement could be caused by this force?



Exercise:

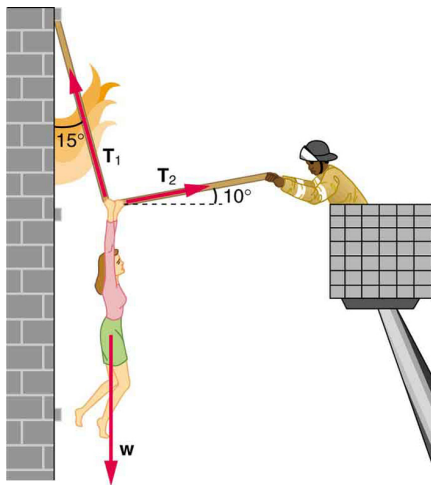
Problem:

A 76.0-kg person is being pulled away from a burning building as shown in [\[link\]](#). Calculate the tension in the two ropes if the person is momentarily motionless. Include a free-body diagram in your solution.

Solution:

$$T_1 = 736 \text{ N}$$

$$T_2 = 194 \text{ N}$$



The force T_2 needed to hold steady the person being rescued from the fire is less than her weight and less than the force T_1 in the other rope, since the more

vertical rope supports a greater part of her weight (a vertical force).

Exercise:

Problem:

Integrated Concepts A 35.0-kg dolphin decelerates from 12.0 to 7.50 m/s in 2.30 s to join another dolphin in play. What average force was exerted to slow him if he was moving horizontally? (The gravitational force is balanced by the buoyant force of the water.)

Exercise:

Problem:

Integrated Concepts When starting a foot race, a 70.0-kg sprinter exerts an average force of 650 N backward on the ground for 0.800 s. (a) What is his final speed? (b) How far does he travel?

Solution:

(a) 7.43 m/s

(b) 2.97 m

Exercise:

Problem:

Integrated Concepts A large rocket has a mass of 2.00×10^6 kg at takeoff, and its engines produce a thrust of 3.50×10^7 N. (a) Find its initial acceleration if it takes off vertically. (b) How long does it take to reach a velocity of 120 km/h straight up, assuming constant mass and thrust? (c) In reality, the mass of a rocket decreases significantly as its fuel is consumed. Describe qualitatively how this affects the acceleration and time for this motion.

Exercise:

Problem:

Integrated Concepts A basketball player jumps straight up for a ball. To do this, he lowers his body 0.300 m and then accelerates through this distance by forcefully straightening his legs. This player leaves the floor with a vertical velocity sufficient to carry him 0.900 m above the floor. (a) Calculate his velocity when he leaves the floor. (b) Calculate his acceleration while he is straightening his legs. He goes from zero to the velocity found in part (a) in a distance of 0.300 m. (c) Calculate the force he exerts on the floor to do this, given that his mass is 110 kg.

Solution:

(a) 4.20 m/s

(b) 29.4 m/s²

(c) 4.31×10^3 N

Exercise:**Problem:**

Integrated Concepts A 2.50-kg fireworks shell is fired straight up from a mortar and reaches a height of 110 m. (a) Neglecting air resistance (a poor assumption, but we will make it for this example), calculate the shell's velocity when it leaves the mortar. (b) The mortar itself is a tube 0.450 m long. Calculate the average acceleration of the shell in the tube as it goes from zero to the velocity found in (a). (c) What is the average force on the shell in the mortar? Express your answer in newtons and as a ratio to the weight of the shell.

Exercise:**Problem:**

Integrated Concepts Repeat [\[link\]](#) for a shell fired at an angle 10.0° from the vertical.

Solution:

(a) 47.1 m/s

(b) 2.47×10^3 m/s²

(c) 6.18×10^3 N . The average force is 252 times the shell's weight.

Exercise:

Problem:

Integrated Concepts An elevator filled with passengers has a mass of 1700 kg. (a) The elevator accelerates upward from rest at a rate of 1.20 m/s^2 for 1.50 s. Calculate the tension in the cable supporting the elevator. (b) The elevator continues upward at constant velocity for 8.50 s. What is the tension in the cable during this time? (c) The elevator decelerates at a rate of 0.600 m/s^2 for 3.00 s. What is the tension in the cable during deceleration? (d) How high has the elevator moved above its original starting point, and what is its final velocity?

Exercise:

Problem:

Unreasonable Results (a) What is the final velocity of a car originally traveling at 50.0 km/h that decelerates at a rate of 0.400 m/s^2 for 50.0 s? (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Exercise:

Problem:

Unreasonable Results A 75.0-kg man stands on a bathroom scale in an elevator that accelerates from rest to 30.0 m/s in 2.00 s. (a) Calculate the scale reading in newtons and compare it with his weight. (The scale exerts an upward force on him equal to its reading.) (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Extended Topic: The Four Basic Forces—An Introduction

- Understand the four basic forces that underlie the processes in nature.

One of the most remarkable simplifications in physics is that only four distinct forces account for all known phenomena. In fact, nearly all of the forces we experience directly are due to only one basic force, called the electromagnetic force. (The gravitational force is the only force we experience directly that is not electromagnetic.) This is a tremendous simplification of the myriad of *apparently* different forces we can list, only a few of which were discussed in the previous section. As we will see, the basic forces are all thought to act through the exchange of microscopic carrier particles, and the characteristics of the basic forces are determined by the types of particles exchanged. Action at a distance, such as the gravitational force of Earth on the Moon, is explained by the existence of a **force field** rather than by “physical contact.”

The *four basic forces* are the gravitational force, the electromagnetic force, the weak nuclear force, and the strong nuclear force. Their properties are summarized in [\[link\]](#). Since the weak and strong nuclear forces act over an extremely short range, the size of a nucleus or less, we do not experience them directly, although they are crucial to the very structure of matter. These forces determine which nuclei are stable and which decay, and they are the basis of the release of energy in certain nuclear reactions. Nuclear forces determine not only the stability of nuclei, but also the relative abundance of elements in nature. The properties of the nucleus of an atom determine the number of electrons it has and, thus, indirectly determine the chemistry of the atom. More will be said of all of these topics in later chapters.

Note:
Concept Connections: The Four Basic Forces

The four basic forces will be encountered in more detail as you progress through the text. The gravitational force is defined in [Uniform Circular Motion and Gravitation](#), electric force in [Electric Charge and Electric Field](#), magnetic force in [Magnetism](#), and nuclear forces in [Radioactivity and Nuclear Physics](#). On a macroscopic scale, electromagnetism and gravity are the basis for all forces. The nuclear forces are vital to the substructure of matter, but they are not directly experienced on the macroscopic scale.

Force	Approximate Relative Strengths	Range	Attraction/Repulsion	Carrier Particle
Gravitational	10^{-38}	∞	attractive only	Graviton

Force	Approximate Relative Strengths	Range	Attraction/Repulsion	Carrier Particle
Electromagnetic	10^{-2}	∞	attractive and repulsive	Photon
Weak nuclear	10^{-13}	$< 10^{-18}\text{m}$	attractive and repulsive	W^+ , W^- , Z^0
Strong nuclear	1	$< 10^{-15}\text{m}$	attractive and repulsive	gluons

Properties of the Four Basic Forces^[footnote]

The graviton is a proposed particle, though it has not yet been observed by scientists. See the discussion of gravitational waves later in this section. The particles W^+ , W^- , and Z^0 are called vector bosons; these were predicted by theory and first observed in 1983. There are eight types of gluons proposed by scientists, and their existence is indicated by meson exchange in the nuclei of atoms.

The gravitational force is surprisingly weak—it is only because gravity is always attractive that we notice it at all. Our weight is the gravitational force due to the *entire* Earth acting on us. On the very large scale, as in astronomical systems, the gravitational force is the dominant force determining the motions of moons, planets, stars, and galaxies. The gravitational force also affects the nature of space and time. As we shall see later in the study of general relativity, space is curved in the vicinity of very massive bodies, such as the Sun, and time actually slows down near massive bodies.

Electromagnetic forces can be either attractive or repulsive. They are long-range forces, which act over extremely large distances, and they nearly cancel for macroscopic objects. (Remember that it is the *net* external force that is important.) If they did not cancel, electromagnetic forces would completely overwhelm the gravitational force. The electromagnetic force is a combination of electrical forces (such as those that cause static electricity) and magnetic forces (such as those that affect a compass needle). These two forces were thought to be quite distinct until early in the 19th century, when scientists began to discover that they are different manifestations of the same force. This discovery is a classical case of the *unification of forces*. Similarly, friction, tension, and all of the other classes of forces we experience directly (except gravity, of course) are due to electromagnetic interactions of atoms and molecules. It is still convenient to consider these forces separately in specific applications, however, because of the ways they manifest themselves.

Note:

Concept Connections: Unifying Forces

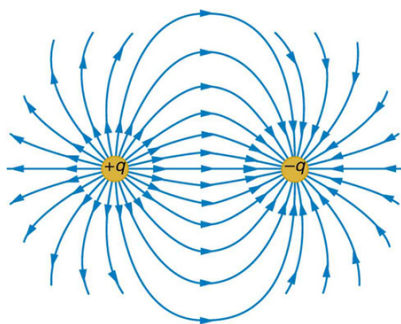
Attempts to unify the four basic forces are discussed in relation to elementary particles later in this text. By “unify” we mean finding connections between the forces that show that they are different manifestations of a single force. Even if such unification is achieved, the forces will retain their separate characteristics on the macroscopic scale and may be identical only under extreme conditions such as those existing in the early universe.

Physicists are now exploring whether the four basic forces are in some way related. Attempts to unify all forces into one come under the rubric of Grand Unified Theories (GUTs), with which there has been some success in recent years. It is now known that under conditions of extremely high density and temperature, such as existed in the early universe, the electromagnetic and weak nuclear forces are indistinguishable. They can now be considered to be different manifestations of one force, called the *electroweak* force. So the list of four has been reduced in a sense to only three. Further progress in unifying all forces is proving difficult—especially the inclusion of the gravitational force, which has the special characteristics of affecting the space and time in which the other forces exist.

While the unification of forces will not affect how we discuss forces in this text, it is fascinating that such underlying simplicity exists in the face of the overt complexity of the universe. There is no reason that nature must be simple—it simply is.

Action at a Distance: Concept of a Field

All forces act at a distance. This is obvious for the gravitational force. Earth and the Moon, for example, interact without coming into contact. It is also true for all other forces. Friction, for example, is an electromagnetic force between atoms that may not actually touch. What is it that carries forces between objects? One way to answer this question is to imagine that a **force field** surrounds whatever object creates the force. A second object (often called a *test object*) placed in this field will experience a force that is a function of location and other variables. The field itself is the “thing” that carries the force from one object to another. The field is defined so as to be a characteristic of the object creating it; the field does not depend on the test object placed in it. Earth’s gravitational field, for example, is a function of the mass of Earth and the distance from its center, independent of the presence of other masses. The concept of a field is useful because equations can be written for force fields surrounding objects (for gravity, this yields $w = mg$ at Earth’s surface), and motions can be calculated from these equations. (See [\[link\]](#).)



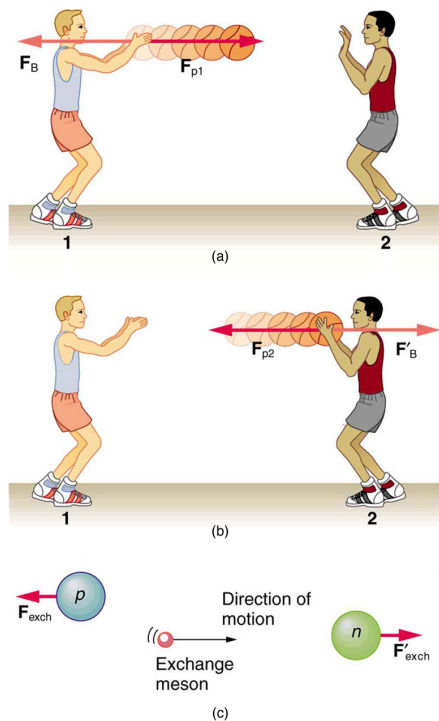
The electric force field between a positively charged particle and a negatively charged particle. When a positive test charge is placed in the field, the charge will experience a force in the direction of the force field lines.

Note:

Concept Connections: Force Fields

The concept of a *force field* is also used in connection with electric charge and is presented in [Electric Charge and Electric Field](#). It is also a useful idea for all the basic forces, as will be seen in [Particle Physics](#). Fields help us to visualize forces and how they are transmitted, as well as to describe them with precision and to link forces with subatomic carrier particles.

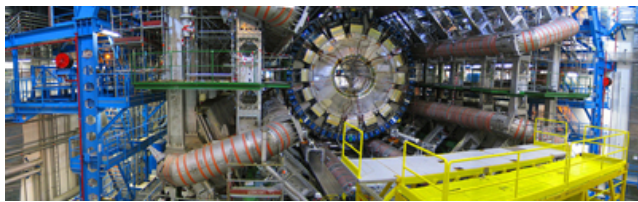
The field concept has been applied very successfully; we can calculate motions and describe nature to high precision using field equations. As useful as the field concept is, however, it leaves unanswered the question of what carries the force. It has been proposed in recent decades, starting in 1935 with Hideki Yukawa's (1907–1981) work on the strong nuclear force, that all forces are transmitted by the exchange of elementary particles. We can visualize particle exchange as analogous to macroscopic phenomena such as two people passing a basketball back and forth, thereby exerting a repulsive force without touching one another. (See [link](#).)



The exchange of masses resulting in repulsive forces.

(a) The person throwing the basketball exerts a force \mathbf{F}_{p1} on it toward the other person and feels a reaction force \mathbf{F}_B away from the second person. (b) The person catching the basketball exerts a force \mathbf{F}_{p2} on it to stop the ball and feels a reaction force \mathbf{F}'_B away from the first person. (c) The analogous exchange of a meson between a proton and a neutron carries the strong nuclear forces \mathbf{F}_{exch} and $\mathbf{F}'_{\text{exch}}$ between them. An attractive force can also be exerted by the exchange of a mass—if person 2 pulled the basketball away from the first person as he tried to retain it, then the force between them would be attractive.

This idea of particle exchange deepens rather than contradicts field concepts. It is more satisfying philosophically to think of something physical actually moving between objects acting at a distance. [\[link\]](#) lists the exchange or **carrier particles**, both observed and proposed, that carry the four forces. But the real fruit of the particle-exchange proposal is that searches for Yukawa's proposed particle found it *and* a number of others that were completely unexpected, stimulating yet more research. All of this research eventually led to the proposal of quarks as the underlying substructure of matter, which is a basic tenet of GUTs. If successful, these theories will explain not only forces, but also the structure of matter itself. Yet physics is an experimental science, so the test of these theories must lie in the domain of the real world. As of this writing, scientists at the CERN laboratory in Switzerland are starting to test these theories using the world's largest particle accelerator: the Large Hadron Collider. This accelerator (27 km in circumference) allows two high-energy proton beams, traveling in opposite directions, to collide. An energy of 14 trillion electron volts will be available. It is anticipated that some new particles, possibly force carrier particles, will be found. (See [\[link\]](#).) One of the force carriers of high interest that researchers hope to detect is the Higgs boson. The observation of its properties might tell us why different particles have different masses.



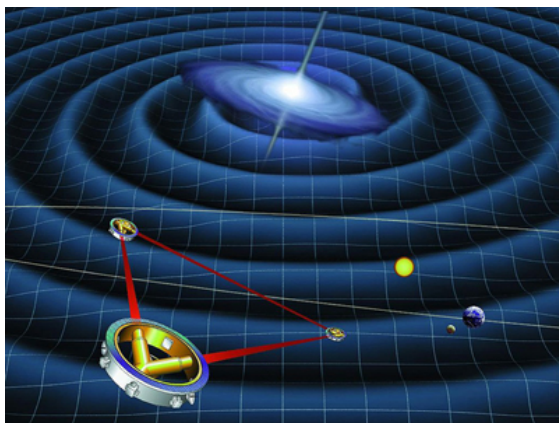
The world's largest particle accelerator spans the border between Switzerland and France. Two beams, traveling in opposite directions close to the speed of light, collide in a tube similar to the central tube shown here. External magnets determine the beam's path. Special detectors will analyze particles created in these collisions. Questions as broad as what is the origin of mass and what was matter like the first few seconds of our universe will be explored. This accelerator began preliminary operation in 2008. (credit: Frank Hommes)

Tiny particles also have wave-like behavior, something we will explore more in a later chapter. To better understand force-carrier particles from another perspective, let us consider gravity. The search for gravitational waves has been going on for a number of years. Almost 100 years

ago, Einstein predicted the existence of these waves as part of his general theory of relativity. Gravitational waves are created during the collision of massive stars, in black holes, or in supernova explosions—like shock waves. These gravitational waves will travel through space from such sites much like a pebble dropped into a pond sends out ripples—except these waves move at the speed of light. A detector apparatus has been built in the U.S., consisting of two large installations nearly 3000 km apart—one in Washington state and one in Louisiana! The facility is called the Laser Interferometer Gravitational-Wave Observatory (LIGO). Each installation is designed to use optical lasers to examine any slight shift in the relative positions of two masses due to the effect of gravity waves. The two sites allow simultaneous measurements of these small effects to be separated from other natural phenomena, such as earthquakes. Initial operation of the detectors began in 2002, and work is proceeding on increasing their sensitivity. Similar installations have been built in Italy (VIRGO), Germany (GEO600), and Japan (TAMA300) to provide a worldwide network of gravitational wave detectors.

International collaboration in this area is moving into space with the joint EU/US project LISA (Laser Interferometer Space Antenna). Earthquakes and other Earthly noises will be no problem for these monitoring spacecraft. LISA will complement LIGO by looking at much more massive black holes through the observation of gravitational-wave sources emitting much larger wavelengths. Three satellites will be placed in space above Earth in an equilateral triangle (with 5,000,000-km sides) ([\[link\]](#)). The system will measure the relative positions of each satellite to detect passing gravitational waves. Accuracy to within 10% of the size of an atom will be needed to detect any waves. The launch of this project might be as early as 2018.

“I’m sure LIGO will tell us something about the universe that we didn’t know before. The history of science tells us that any time you go where you haven’t been before, you usually find something that really shakes the scientific paradigms of the day. Whether gravitational wave astrophysics will do that, only time will tell.” —David Reitze, LIGO Input Optics Manager, University of Florida



Space-based future experiments for the measurement of gravitational waves. Shown here is a drawing of

LISA's orbit. Each satellite of LISA will consist of a laser source and a mass. The lasers will transmit a signal to measure the distance between each satellite's test mass. The relative motion of these masses will provide information about passing gravitational waves. (credit: NASA)

The ideas presented in this section are but a glimpse into topics of modern physics that will be covered in much greater depth in later chapters.

Summary

- The various types of forces that are categorized for use in many applications are all manifestations of the *four basic forces* in nature.
- The properties of these forces are summarized in [\[link\]](#).
- Everything we experience directly without sensitive instruments is due to either electromagnetic forces or gravitational forces. The nuclear forces are responsible for the submicroscopic structure of matter, but they are not directly sensed because of their short ranges. Attempts are being made to show all four forces are different manifestations of a single unified force.
- A force field surrounds an object creating a force and is the carrier of that force.

Conceptual Questions

Exercise:

Problem:

Explain, in terms of the properties of the four basic forces, why people notice the gravitational force acting on their bodies if it is such a comparatively weak force.

Exercise:

Problem:

What is the dominant force between astronomical objects? Why are the other three basic forces less significant over these very large distances?

Exercise:

Problem:

Give a detailed example of how the exchange of a particle can result in an *attractive* force. (For example, consider one child pulling a toy out of the hands of another.)

Problem Exercises

Exercise:

Problem:

(a) What is the strength of the weak nuclear force relative to the strong nuclear force? (b) What is the strength of the weak nuclear force relative to the electromagnetic force? Since the weak nuclear force acts at only very short distances, such as inside nuclei, where the strong and electromagnetic forces also act, it might seem surprising that we have any knowledge of it at all. We have such knowledge because the weak nuclear force is responsible for beta decay, a type of nuclear decay not explained by other forces.

Solution:

(a) 1×10^{-13}

(b) 1×10^{-11}

Exercise:

Problem:

(a) What is the ratio of the strength of the gravitational force to that of the strong nuclear force? (b) What is the ratio of the strength of the gravitational force to that of the weak nuclear force? (c) What is the ratio of the strength of the gravitational force to that of the electromagnetic force? What do your answers imply about the influence of the gravitational force on atomic nuclei?

Exercise:

Problem:

What is the ratio of the strength of the strong nuclear force to that of the electromagnetic force? Based on this ratio, you might expect that the strong force dominates the nucleus, which is true for small nuclei. Large nuclei, however, have sizes greater than the range of the strong nuclear force. At these sizes, the electromagnetic force begins to affect nuclear stability. These facts will be used to explain nuclear fusion and fission later in this text.

Solution:

10^2

Glossary

carrier particle

a fundamental particle of nature that is surrounded by a characteristic force field; photons are carrier particles of the electromagnetic force

force field

a region in which a test particle will experience a force

Introduction: Further Applications of Newton's Laws

class="introduction"

Total hip
replacemen
t surgery
has become
a common
procedure.

The head
(or ball) of
the
patient's
femur fits
into a cup
that has a
hard
plastic-like
inner
lining.

(credit:
National
Institutes of
Health, via
Wikimedia
Commons)



Describe the forces on the hip joint. What means are taken to ensure that this will be a good movable joint? From the photograph (for an adult) in [\[link\]](#), estimate the dimensions of the artificial device.

It is difficult to categorize forces into various types (aside from the four basic forces discussed in previous chapter). We know that a net force affects the motion, position, and shape of an object. It is useful at this point to look at some particularly interesting and common forces that will provide further applications of Newton's laws of motion. We have in mind the forces of friction, air or liquid drag, and deformation.

Friction

- Discuss the general characteristics of friction.
- Describe the various types of friction.
- Calculate the magnitude of static and kinetic friction.

Friction is a force that is around us all the time that opposes relative motion between systems in contact but also allows us to move (which you have discovered if you have ever tried to walk on ice). While a common force, the behavior of friction is actually very complicated and is still not completely understood. We have to rely heavily on observations for whatever understandings we can gain. However, we can still deal with its more elementary general characteristics and understand the circumstances in which it behaves.

Note:

Friction

Friction is a force that opposes relative motion between systems in contact.

One of the simpler characteristics of friction is that it is parallel to the contact surface between systems and always in a direction that opposes motion or attempted motion of the systems relative to each other. If two systems are in contact and moving relative to one another, then the friction between them is called **kinetic friction**. For example, friction slows a hockey puck sliding on ice. But when objects are stationary, **static friction** can act between them; the static friction is usually greater than the kinetic friction between the objects.

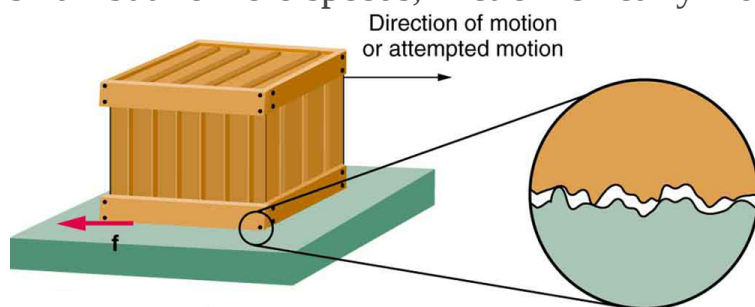
Note:

Kinetic Friction

If two systems are in contact and moving relative to one another, then the friction between them is called kinetic friction.

Imagine, for example, trying to slide a heavy crate across a concrete floor—you may push harder and harder on the crate and not move it at all. This means that the static friction responds to what you do—it increases to be equal to and in the opposite direction of your push. But if you finally push hard enough, the crate seems to slip suddenly and starts to move. Once in motion it is easier to keep it in motion than it was to get it started, indicating that the kinetic friction force is less than the static friction force. If you add mass to the crate, say by placing a box on top of it, you need to push even harder to get it started and also to keep it moving. Furthermore, if you oiled the concrete you would find it to be easier to get the crate started and keep it going (as you might expect).

[\[link\]](#) is a crude pictorial representation of how friction occurs at the interface between two objects. Close-up inspection of these surfaces shows them to be rough. So when you push to get an object moving (in this case, a crate), you must raise the object until it can skip along with just the tips of the surface hitting, break off the points, or do both. A considerable force can be resisted by friction with no apparent motion. The harder the surfaces are pushed together (such as if another box is placed on the crate), the more force is needed to move them. Part of the friction is due to adhesive forces between the surface molecules of the two objects, which explain the dependence of friction on the nature of the substances. Adhesion varies with substances in contact and is a complicated aspect of surface physics. Once an object is moving, there are fewer points of contact (fewer molecules adhering), so less force is required to keep the object moving. At small but nonzero speeds, friction is nearly independent of speed.



Frictional forces, such as f , always oppose motion or attempted motion between objects in contact. Friction arises in part because of the roughness of

the surfaces in contact, as seen in the expanded view. In order for the object to move, it must rise to where the peaks can skip along the bottom surface. Thus a force is required just to set the object in motion. Some of the peaks will be broken off, also requiring a force to maintain motion. Much of the friction is actually due to attractive forces between molecules making up the two objects, so that even perfectly smooth surfaces are not friction-free. Such adhesive forces also depend on the substances the surfaces are made of, explaining, for example, why rubber-soled shoes slip less than those with leather soles.

The magnitude of the frictional force has two forms: one for static situations (static friction), the other for when there is motion (kinetic friction).

When there is no motion between the objects, the **magnitude of static friction** f_s is

Equation:

$$f_s \leq \mu_s N,$$

where μ_s is the coefficient of static friction and N is the magnitude of the normal force (the force perpendicular to the surface).

Note:

Magnitude of Static Friction

Magnitude of static friction f_s is

Equation:

$$f_s \leq \mu_s N,$$

where μ_s is the coefficient of static friction and N is the magnitude of the normal force.

The symbol \leq means *less than or equal to*, implying that static friction can have a minimum and a maximum value of $\mu_s N$. Static friction is a responsive force that increases to be equal and opposite to whatever force is exerted, up to its maximum limit. Once the applied force exceeds $f_{s(\max)}$, the object will move. Thus

Equation:

$$f_{s(\max)} = \mu_s N.$$

Once an object is moving, the **magnitude of kinetic friction** f_k is given by **Equation:**

$$f_k = \mu_k N,$$

where μ_k is the coefficient of kinetic friction. A system in which $f_k = \mu_k N$ is described as a system in which *friction behaves simply*.

Note:**Magnitude of Kinetic Friction**

The magnitude of kinetic friction f_k is given by

Equation:

$$f_k = \mu_k N,$$

where μ_k is the coefficient of kinetic friction.

As seen in [\[link\]](#), the coefficients of kinetic friction are less than their static counterparts. That values of μ in [\[link\]](#) are stated to only one or, at most, two digits is an indication of the approximate description of friction given by the above two equations.

System	Static friction μ_s	Kinetic friction μ_k
Rubber on dry concrete	1.0	0.7
Rubber on wet concrete	0.7	0.5
Wood on wood	0.5	0.3
Waxed wood on wet snow	0.14	0.1
Metal on wood	0.5	0.3
Steel on steel (dry)	0.6	0.3
Steel on steel (oiled)	0.05	0.03
Teflon on steel	0.04	0.04
Bone lubricated by synovial fluid	0.016	0.015
Shoes on wood	0.9	0.7

System	Static friction μ_s	Kinetic friction μ_k
Shoes on ice	0.1	0.05
Ice on ice	0.1	0.03
Steel on ice	0.04	0.02

Coefficients of Static and Kinetic Friction

The equations given earlier include the dependence of friction on materials and the normal force. The direction of friction is always opposite that of motion, parallel to the surface between objects, and perpendicular to the normal force. For example, if the crate you try to push (with a force parallel to the floor) has a mass of 100 kg, then the normal force would be equal to its weight, $W = mg = (100 \text{ kg})(9.80 \text{ m/s}^2) = 980 \text{ N}$, perpendicular to the floor. If the coefficient of static friction is 0.45, you would have to exert a force parallel to the floor greater than

$f_{s(\text{max})} = \mu_s N = (0.45)(980 \text{ N}) = 440 \text{ N}$ to move the crate. Once there is motion, friction is less and the coefficient of kinetic friction might be 0.30, so that a force of only 290 N ($f_k = \mu_k N = (0.30)(980 \text{ N}) = 290 \text{ N}$) would keep it moving at a constant speed. If the floor is lubricated, both coefficients are considerably less than they would be without lubrication. Coefficient of friction is a unit less quantity with a magnitude usually between 0 and 1.0. The coefficient of the friction depends on the two surfaces that are in contact.

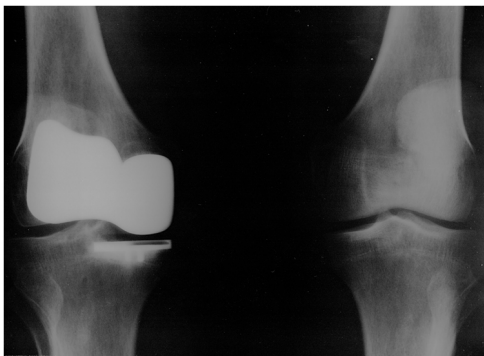
Note:

Take-Home Experiment

Find a small plastic object (such as a food container) and slide it on a kitchen table by giving it a gentle tap. Now spray water on the table,

simulating a light shower of rain. What happens now when you give the object the same-sized tap? Now add a few drops of (vegetable or olive) oil on the surface of the water and give the same tap. What happens now? This latter situation is particularly important for drivers to note, especially after a light rain shower. Why?

Many people have experienced the slipperiness of walking on ice. However, many parts of the body, especially the joints, have much smaller coefficients of friction—often three or four times less than ice. A joint is formed by the ends of two bones, which are connected by thick tissues. The knee joint is formed by the lower leg bone (the tibia) and the thighbone (the femur). The hip is a ball (at the end of the femur) and socket (part of the pelvis) joint. The ends of the bones in the joint are covered by cartilage, which provides a smooth, almost glassy surface. The joints also produce a fluid (synovial fluid) that reduces friction and wear. A damaged or arthritic joint can be replaced by an artificial joint ([link](#)). These replacements can be made of metals (stainless steel or titanium) or plastic (polyethylene), also with very small coefficients of friction.



Artificial knee replacement is a procedure that has been performed for more than 20 years. In this figure, we see the post-op x rays of the right knee joint replacement. (credit: Mike Baird, Flickr)

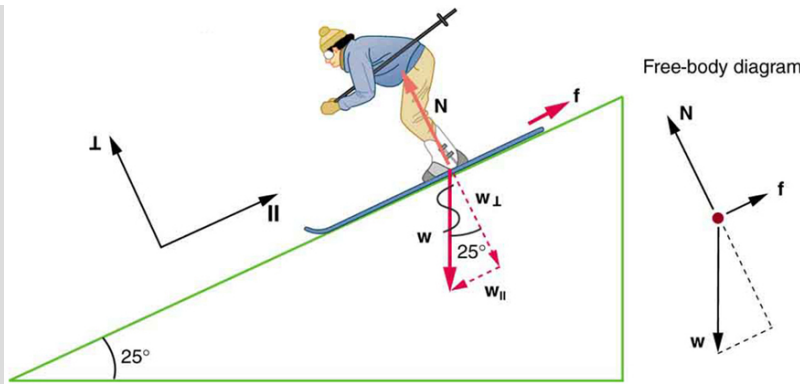
Other natural lubricants include saliva produced in our mouths to aid in the swallowing process, and the slippery mucus found between organs in the body, allowing them to move freely past each other during heartbeats, during breathing, and when a person moves. Artificial lubricants are also common in hospitals and doctor's clinics. For example, when ultrasonic imaging is carried out, the gel that couples the transducer to the skin also serves to lubricate the surface between the transducer and the skin—thereby reducing the coefficient of friction between the two surfaces. This allows the transducer to move freely over the skin.

Example:**Skiing Exercise**

A skier with a mass of 62 kg is sliding down a snowy slope. Find the coefficient of kinetic friction for the skier if friction is known to be 45.0 N.

Strategy

The magnitude of kinetic friction was given in to be 45.0 N. Kinetic friction is related to the normal force N as $f_k = \mu_k N$; thus, the coefficient of kinetic friction can be found if we can find the normal force of the skier on a slope. The normal force is always perpendicular to the surface, and since there is no motion perpendicular to the surface, the normal force should equal the component of the skier's weight perpendicular to the slope. (See the skier and free-body diagram in [\[link\]](#).)



The motion of the skier and friction are parallel to the slope and so it is most convenient to project all forces onto a coordinate system where one axis is parallel to the slope and the other is perpendicular (axes shown to left of skier). \mathbf{N} (the normal force) is perpendicular to the slope, and \mathbf{f} (the friction) is parallel to the slope, but \mathbf{w} (the skier's weight) has components along both axes, namely \mathbf{w}_{\perp} and \mathbf{W}_{\parallel} . \mathbf{N} is equal in magnitude to \mathbf{w}_{\perp} , so there is no motion perpendicular to the slope. However, \mathbf{f} is less than \mathbf{W}_{\parallel} in magnitude, so there is acceleration down the slope (along the x -axis).

That is,

Equation:

$$N = w_{\perp} = w \cos 25^{\circ} = mg \cos 25^{\circ}.$$

Substituting this into our expression for kinetic friction, we get

Equation:

$$f_k = \mu_k mg \cos 25^{\circ},$$

which can now be solved for the coefficient of kinetic friction μ_k .

Solution

Solving for μ_k gives

Equation:

$$\mu_k = \frac{f_k}{N} = \frac{f_k}{w \cos 25^\circ} = \frac{f_k}{mg \cos 25^\circ}.$$

Substituting known values on the right-hand side of the equation,

Equation:

$$\mu_k = \frac{45.0 \text{ N}}{(62 \text{ kg})(9.80 \text{ m/s}^2)(0.906)} = 0.082.$$

Discussion

This result is a little smaller than the coefficient listed in [\[link\]](#) for waxed wood on snow, but it is still reasonable since values of the coefficients of friction can vary greatly. In situations like this, where an object of mass m slides down a slope that makes an angle θ with the horizontal, friction is given by $f_k = \mu_k mg \cos \theta$. All objects will slide down a slope with constant acceleration under these circumstances. Proof of this is left for this chapter's Problems and Exercises.

Note:**Take-Home Experiment**

An object will slide down an inclined plane at a constant velocity if the net force on the object is zero. We can use this fact to measure the coefficient of kinetic friction between two objects. As shown in [\[link\]](#), the kinetic friction on a slope $f_k = \mu_k mg \cos \theta$. The component of the weight down the slope is equal to $mg \sin \theta$ (see the free-body diagram in [\[link\]](#)). These forces act in opposite directions, so when they have equal magnitude, the acceleration is zero. Writing these out:

Equation:

$$f_k = F_{g_x}$$

Equation:

$$\mu_k mg \cos \theta = mg \sin \theta.$$

Solving for μ_k , we find that

Equation:

$$\mu_k = \frac{mg \sin \theta}{mg \cos \theta} = \tan \theta.$$

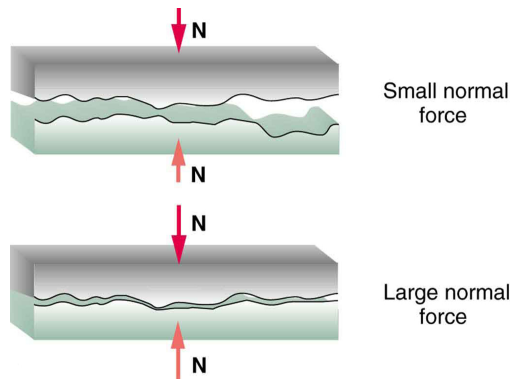
Put a coin on a book and tilt it until the coin slides at a constant velocity down the book. You might need to tap the book lightly to get the coin to move. Measure the angle of tilt relative to the horizontal and find μ_k . Note that the coin will not start to slide at all until an angle greater than θ is attained, since the coefficient of static friction is larger than the coefficient of kinetic friction. Discuss how this may affect the value for μ_k and its uncertainty.

We have discussed that when an object rests on a horizontal surface, there is a normal force supporting it equal in magnitude to its weight. Furthermore, simple friction is always proportional to the normal force.

Note:**Making Connections: Submicroscopic Explanations of Friction**

The simpler aspects of friction dealt with so far are its macroscopic (large-scale) characteristics. Great strides have been made in the atomic-scale explanation of friction during the past several decades. Researchers are finding that the atomic nature of friction seems to have several fundamental characteristics. These characteristics not only explain some of the simpler aspects of friction—they also hold the potential for the development of nearly friction-free environments that could save hundreds of billions of dollars in energy which is currently being converted (unnecessarily) to heat.

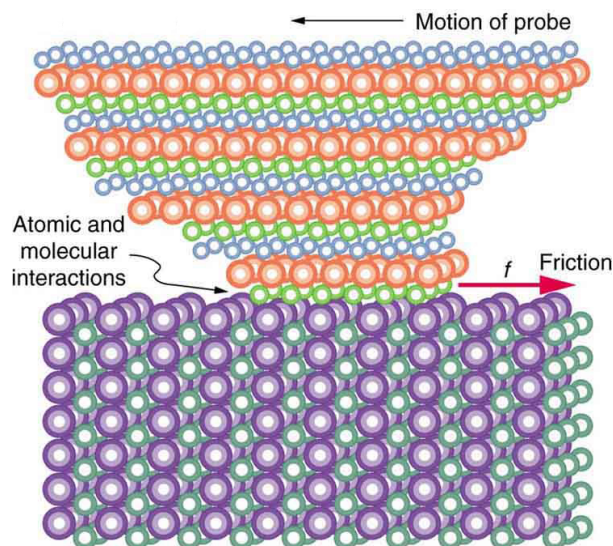
[\[link\]](#) illustrates one macroscopic characteristic of friction that is explained by microscopic (small-scale) research. We have noted that friction is proportional to the normal force, but not to the area in contact, a somewhat counterintuitive notion. When two rough surfaces are in contact, the actual contact area is a tiny fraction of the total area since only high spots touch. When a greater normal force is exerted, the actual contact area increases, and it is found that the friction is proportional to this area.



Two rough surfaces in contact have a much smaller area of actual contact than their total area. When there is a greater normal force as a result of a greater applied force, the area of actual contact increases as does friction.

But the atomic-scale view promises to explain far more than the simpler features of friction. The mechanism for how heat is generated is now being determined. In other words, why do surfaces get warmer when rubbed? Essentially, atoms are linked with one another to form lattices. When surfaces rub, the surface atoms adhere and cause atomic lattices to vibrate—essentially creating sound waves that penetrate the material. The sound waves diminish with distance and their energy is converted into heat. Chemical reactions that are related to frictional wear can also occur

between atoms and molecules on the surfaces. [\[link\]](#) shows how the tip of a probe drawn across another material is deformed by atomic-scale friction. The force needed to drag the tip can be measured and is found to be related to shear stress, which will be discussed later in this chapter. The variation in shear stress is remarkable (more than a factor of 10^{12}) and difficult to predict theoretically, but shear stress is yielding a fundamental understanding of a large-scale phenomenon known since ancient times—friction.



The tip of a probe is deformed sideways by frictional force as the probe is dragged across a surface. Measurements of how the force varies for different materials are yielding fundamental insights into the atomic nature of friction.

Note:

PhET Explorations: Forces and Motion

Explore the forces at work when you try to push a filing cabinet. Create an applied force and see the resulting friction force and total force acting on the cabinet. Charts show the forces, position, velocity, and acceleration vs. time. Draw a free-body diagram of all the forces (including gravitational and normal forces).

[Forces](#)
[and](#)
[Motion](#)
[Simulation](#)

Section Summary

- Friction is a contact force between systems that opposes the motion or attempted motion between them. Simple friction is proportional to the normal force N pushing the systems together. (A normal force is always perpendicular to the contact surface between systems.) Friction depends on both of the materials involved. The magnitude of static friction f_s between systems stationary relative to one another is given by

Equation:

$$f_s \leq \mu_s N,$$

where μ_s is the coefficient of static friction, which depends on both of the materials.

- The kinetic friction force f_k between systems moving relative to one another is given by

Equation:

$$f_k = \mu_k N,$$

where μ_k is the coefficient of kinetic friction, which also depends on both materials.

Conceptual Questions

Exercise:

Problem:

Define normal force. What is its relationship to friction when friction behaves simply?

Exercise:

Problem:

The glue on a piece of tape can exert forces. Can these forces be a type of simple friction? Explain, considering especially that tape can stick to vertical walls and even to ceilings.

Exercise:

Problem:

When you learn to drive, you discover that you need to let up slightly on the brake pedal as you come to a stop or the car will stop with a jerk. Explain this in terms of the relationship between static and kinetic friction.

Exercise:

Problem:

When you push a piece of chalk across a chalkboard, it sometimes screeches because it rapidly alternates between slipping and sticking to the board. Describe this process in more detail, in particular explaining how it is related to the fact that kinetic friction is less than static friction. (The same slip-grab process occurs when tires screech on pavement.)

Problems & Exercises

Exercise:

Problem:

A physics major is cooking breakfast when he notices that the frictional force between his steel spatula and his Teflon frying pan is only 0.200 N. Knowing the coefficient of kinetic friction between the two materials, he quickly calculates the normal force. What is it?

Solution:

5.00 N

Exercise:

Problem:

(a) When rebuilding her car's engine, a physics major must exert 300 N of force to insert a dry steel piston into a steel cylinder. What is the magnitude of the normal force between the piston and cylinder? (b) What is the magnitude of the force would she have to exert if the steel parts were oiled?

Exercise:

Problem:

(a) What is the maximum frictional force in the knee joint of a person who supports 66.0 kg of her mass on that knee? (b) During strenuous exercise it is possible to exert forces to the joints that are easily ten times greater than the weight being supported. What is the maximum force of friction under such conditions? The frictional forces in joints are relatively small in all circumstances except when the joints deteriorate, such as from injury or arthritis. Increased frictional forces can cause further damage and pain.

Exercise:

Problem:

Suppose you have a 120-kg wooden crate resting on a wood floor. (a) What maximum force can you exert horizontally on the crate without moving it? (b) If you continue to exert this force once the crate starts to slip, what will the magnitude of its acceleration then be?

Solution:

(a) 588 N

(b) 1.96 m/s^2

Exercise:**Problem:**

(a) If half of the weight of a small $1.00 \times 10^3 \text{ kg}$ utility truck is supported by its two drive wheels, what is the magnitude of the maximum acceleration it can achieve on dry concrete? (b) Will a metal cabinet lying on the wooden bed of the truck slip if it accelerates at this rate? (c) Solve both problems assuming the truck has four-wheel drive.

Exercise:**Problem:**

A team of eight dogs pulls a sled with waxed wood runners on wet snow (mush!). The dogs have average masses of 19.0 kg, and the loaded sled with its rider has a mass of 210 kg. (a) Calculate the magnitude of the acceleration starting from rest if each dog exerts an average force of 185 N backward on the snow. (b) What is the magnitude of the acceleration once the sled starts to move? (c) For both situations, calculate the magnitude of the force in the coupling between the dogs and the sled.

Solution:

(a) 3.29 m/s^2

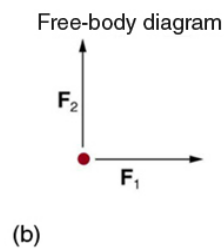
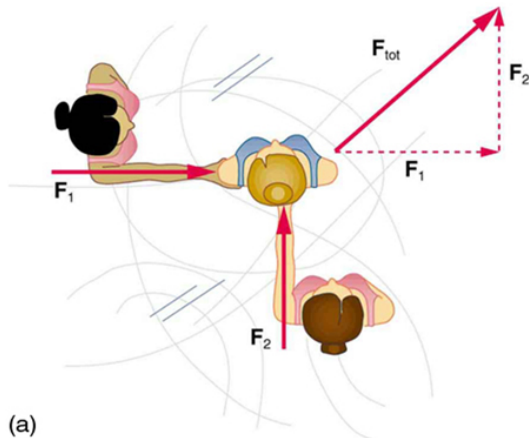
(b) 3.52 m/s^2

(c) 980 N; 945 N

Exercise:

Problem:

Consider the 65.0-kg ice skater being pushed by two others shown in [\[link\]](#). (a) Find the direction and magnitude of \mathbf{F}_{tot} , the total force exerted on her by the others, given that the magnitudes F_1 and F_2 are 26.4 N and 18.6 N, respectively. (b) What is her initial acceleration if she is initially stationary and wearing steel-bladed skates that point in the direction of \mathbf{F}_{tot} ? (c) What is her acceleration assuming she is already moving in the direction of \mathbf{F}_{tot} ? (Remember that friction always acts in the direction opposite that of motion or attempted motion between surfaces in contact.)



Exercise:

Problem:

Show that the acceleration of any object down a frictionless incline that makes an angle θ with the horizontal is $a = g \sin \theta$. (Note that this acceleration is independent of mass.)

Exercise:

Problem:

Show that the acceleration of any object down an incline where friction behaves simply (that is, where $f_k = \mu_k N$) is $a = g(\sin \theta - \mu_k \cos \theta)$. Note that the acceleration is independent of mass and reduces to the expression found in the previous problem when friction becomes negligibly small ($\mu_k = 0$).

Exercise:**Problem:**

Calculate the deceleration of a snow boarder going up a 5.0° slope assuming the coefficient of friction for waxed wood on wet snow. The result of [\[link\]](#) may be useful, but be careful to consider the fact that the snow boarder is going uphill. Explicitly show how you follow the steps in [Problem-Solving Strategies](#).

Solution:

$$1.83 \text{ m/s}^2$$

Exercise:**Problem:**

(a) Calculate the acceleration of a skier heading down a 10.0° slope, assuming the coefficient of friction for waxed wood on wet snow. (b) Find the angle of the slope down which this skier could coast at a constant velocity. You can neglect air resistance in both parts, and you will find the result of [\[link\]](#) to be useful. Explicitly show how you follow the steps in the [Problem-Solving Strategies](#).

Exercise:

Problem:

If an object is to rest on an incline without slipping, then friction must equal the component of the weight of the object parallel to the incline. This requires greater and greater friction for steeper slopes. Show that the maximum angle of an incline above the horizontal for which an object will not slide down is $\theta = \tan^{-1} \mu_s$. You may use the result of the previous problem. Assume that $a = 0$ and that static friction has reached its maximum value.

Exercise:**Problem:**

Calculate the maximum deceleration of a car that is heading down a 6° slope (one that makes an angle of 6° with the horizontal) under the following road conditions. You may assume that the weight of the car is evenly distributed on all four tires and that the coefficient of static friction is involved—that is, the tires are not allowed to slip during the deceleration. (Ignore rolling.) Calculate for a car: (a) On dry concrete. (b) On wet concrete. (c) On ice, assuming that $\mu_s = 0.100$, the same as for shoes on ice.

Exercise:**Problem:**

Calculate the maximum acceleration of a car that is heading up a 4° slope (one that makes an angle of 4° with the horizontal) under the following road conditions. Assume that only half the weight of the car is supported by the two drive wheels and that the coefficient of static friction is involved—that is, the tires are not allowed to slip during the acceleration. (Ignore rolling.) (a) On dry concrete. (b) On wet concrete. (c) On ice, assuming that $\mu_s = 0.100$, the same as for shoes on ice.

Solution:

(a) 4.20 m/s^2

(b) 2.74 m/s^2

(c) -0.195 m/s^2

Exercise:

Problem: Repeat [\[link\]](#) for a car with four-wheel drive.

Exercise:

Problem:

A freight train consists of two 8.00×10^5 -kg engines and 45 cars with average masses of 5.50×10^5 kg. (a) What force must each engine exert backward on the track to accelerate the train at a rate of $5.00 \times 10^{-2} \text{ m/s}^2$ if the force of friction is $7.50 \times 10^5 \text{ N}$, assuming the engines exert identical forces? This is not a large frictional force for such a massive system. Rolling friction for trains is small, and consequently trains are very energy-efficient transportation systems. (b) What is the magnitude of the force in the coupling between the 37th and 38th cars (this is the force each exerts on the other), assuming all cars have the same mass and that friction is evenly distributed among all of the cars and engines?

Solution:

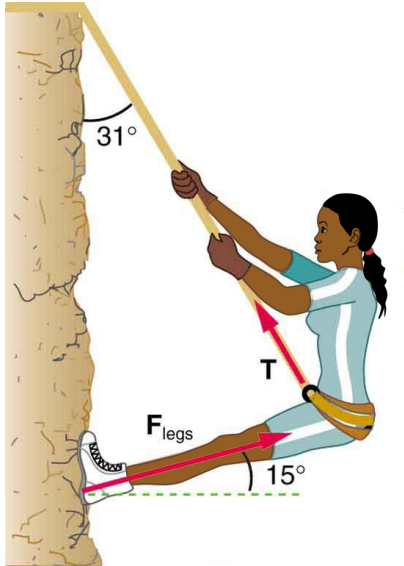
(a) $1.03 \times 10^6 \text{ N}$

(b) $3.48 \times 10^5 \text{ N}$

Exercise:

Problem:

Consider the 52.0-kg mountain climber in [\[link\]](#). (a) Find the tension in the rope and the force that the mountain climber must exert with her feet on the vertical rock face to remain stationary. Assume that the force is exerted parallel to her legs. Also, assume negligible force exerted by her arms. (b) What is the minimum coefficient of friction between her shoes and the cliff?



Part of the climber's weight is supported by her rope and part by friction between her feet and the rock face.

Exercise:

Problem:

A contestant in a winter sporting event pushes a 45.0-kg block of ice across a frozen lake as shown in [\[link\]](#)(a). (a) Calculate the minimum force F he must exert to get the block moving. (b) What is the magnitude of its acceleration once it starts to move, if that force is maintained?

Solution:

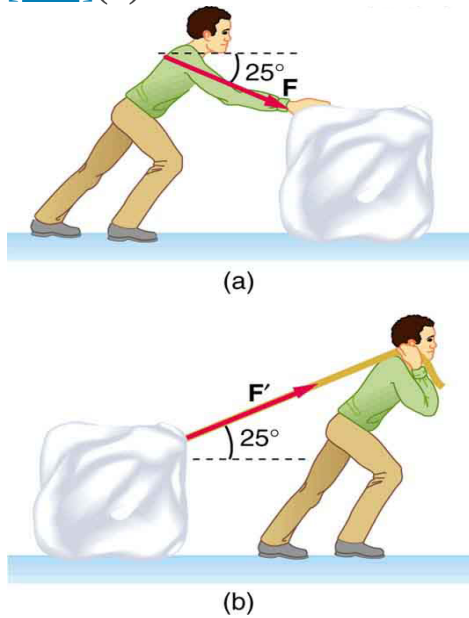
(a) 51.0 N

(b) 0.720 m/s^2

Exercise:

Problem:

Repeat [\[link\]](#) with the contestant pulling the block of ice with a rope over his shoulder at the same angle above the horizontal as shown in [\[link\]](#)(b).



Which method of sliding a block of ice requires less force—(a) pushing or (b) pulling at the same angle above the horizontal?

Glossary

friction

a force that opposes relative motion or attempts at motion between systems in contact

kinetic friction

a force that opposes the motion of two systems that are in contact and moving relative to one another

static friction

a force that opposes the motion of two systems that are in contact and are not moving relative to one another

magnitude of static friction

$f_s \leq \mu_s N$, where μ_s is the coefficient of static friction and N is the magnitude of the normal force

magnitude of kinetic friction

$f_k = \mu_k N$, where μ_k is the coefficient of kinetic friction

Drag Forces

- Express mathematically the drag force.
- Discuss the applications of drag force.
- Define terminal velocity.
- Determine the terminal velocity given mass.

Another interesting force in everyday life is the force of drag on an object when it is moving in a fluid (either a gas or a liquid). You feel the drag force when you move your hand through water. You might also feel it if you move your hand during a strong wind. The faster you move your hand, the harder it is to move. You feel a smaller drag force when you tilt your hand so only the side goes through the air—you have decreased the area of your hand that faces the direction of motion. Like friction, the **drag force** always opposes the motion of an object. Unlike simple friction, the drag force is proportional to some function of the velocity of the object in that fluid. This functionality is complicated and depends upon the shape of the object, its size, its velocity, and the fluid it is in. For most large objects such as bicyclists, cars, and baseballs not moving too slowly, the magnitude of the drag force F_D is found to be proportional to the square of the speed of the object. We can write this relationship mathematically as $F_D \propto v^2$. When taking into account other factors, this relationship becomes

Equation:

$$F_D = \frac{1}{2} C \rho A v^2,$$

where C is the drag coefficient, A is the area of the object facing the fluid, and ρ is the density of the fluid. (Recall that density is mass per unit volume.) This equation can also be written in a more generalized fashion as $F_D = b v^2$, where b is a constant equivalent to $0.5 C \rho A$. We have set the exponent for these equations as 2 because, when an object is moving at high velocity through air, the magnitude of the drag force is proportional to the square of the speed. As we shall see in a few pages on fluid dynamics, for small particles moving at low speeds in a fluid, the exponent is equal to 1.

Note:**Drag Force**

Drag force F_D is found to be proportional to the square of the speed of the object. Mathematically

Equation:

$$F_D \propto v^2$$

Equation:

$$F_D = \frac{1}{2} C \rho A v^2,$$

where C is the drag coefficient, A is the area of the object facing the fluid, and ρ is the density of the fluid.

Athletes as well as car designers seek to reduce the drag force to lower their race times. (See [\[link\]](#)). “Aerodynamic” shaping of an automobile can reduce the drag force and so increase a car’s gas mileage.



From racing cars to bobsled racers, aerodynamic shaping is crucial to achieving top speeds. Bobsleds are designed for speed. They are shaped like a bullet with tapered fins. (credit:

U.S. Army, via
Wikimedia Commons)

The value of the drag coefficient, C , is determined empirically, usually with the use of a wind tunnel. (See [\[link\]](#)).



NASA researchers
test a model plane
in a wind tunnel.
(credit:
NASA/Ames)

The drag coefficient can depend upon velocity, but we will assume that it is a constant here. [\[link\]](#) lists some typical drag coefficients for a variety of objects. Notice that the drag coefficient is a dimensionless quantity. At highway speeds, over 50% of the power of a car is used to overcome air drag. The most fuel-efficient cruising speed is about 70–80 km/h (about 45–50 mi/h). For this reason, during the 1970s oil crisis in the United States, maximum speeds on highways were set at about 90 km/h (55 mi/h).

Object	C
Airfoil	0.05
Toyota Camry	0.28
Ford Focus	0.32
Honda Civic	0.36
Ferrari Testarossa	0.37
Dodge Ram pickup	0.43
Sphere	0.45
Hummer H2 SUV	0.64
Skydiver (feet first)	0.70
Bicycle	0.90
Skydiver (horizontal)	1.0
Circular flat plate	1.12

Drag Coefficient Values Typical values of drag coefficient C .

Substantial research is under way in the sporting world to minimize drag. The dimples on golf balls are being redesigned as are the clothes that athletes wear. Bicycle racers and some swimmers and runners wear full bodysuits. Australian Cathy Freeman wore a full body suit in the 2000 Sydney Olympics, and won the gold medal for the 400 m race. Many swimmers in the 2008 Beijing Olympics wore (Speedo) body suits; it might have made a difference in breaking many world records (See [\[link\]](#)). Most elite swimmers (and cyclists) shave their body hair. Such innovations can have the effect of slicing away milliseconds in a race, sometimes making

the difference between a gold and a silver medal. One consequence is that careful and precise guidelines must be continuously developed to maintain the integrity of the sport.



Body suits, such as this LZR Racer Suit, have been credited with many world records after their release in 2008. Smoother “skin” and more compression forces on a swimmer’s body provide at least 10% less drag. (credit: NASA/Kathy Barnstorff)

Some interesting situations connected to Newton’s second law occur when considering the effects of drag forces upon a moving object. For instance, consider a skydiver falling through air under the influence of gravity. The two forces acting on him are the force of gravity and the drag force (ignoring the buoyant force). The downward force of gravity remains constant regardless of the velocity at which the person is moving. However, as the person’s velocity increases, the magnitude of the drag force increases until the magnitude of the drag force is equal to the gravitational force, thus producing a net force of zero. A zero net force means that there is no

acceleration, as given by Newton's second law. At this point, the person's velocity remains constant and we say that the person has reached his *terminal velocity* (v_t). Since F_D is proportional to the speed, a heavier skydiver must go faster for F_D to equal his weight. Let's see how this works out more quantitatively.

At the terminal velocity,

Equation:

$$F_{\text{net}} = mg - F_D = ma = 0.$$

Thus,

Equation:

$$mg = F_D.$$

Using the equation for drag force, we have

Equation:

$$mg = \frac{1}{2} \rho C A v^2.$$

Solving for the velocity, we obtain

Equation:

$$v = \sqrt{\frac{2mg}{\rho C A}}.$$

Assume the density of air is $\rho = 1.21 \text{ kg/m}^3$. A 75-kg skydiver descending head first will have an area approximately $A = 0.18 \text{ m}^2$ and a drag coefficient of approximately $C = 0.70$. We find that

Equation:

$$\begin{aligned}
 v &= \sqrt{\frac{2(75 \text{ kg})(9.80 \text{ m/s}^2)}{(1.21 \text{ kg/m}^3)(0.70)(0.18 \text{ m}^2)}} \\
 &= 98 \text{ m/s} \\
 &= 350 \text{ km/h.}
 \end{aligned}$$

This means a skydiver with a mass of 75 kg achieves a maximum terminal velocity of about 350 km/h while traveling in a pike (head first) position, minimizing the area and his drag. In a spread-eagle position, that terminal velocity may decrease to about 200 km/h as the area increases. This terminal velocity becomes much smaller after the parachute opens.

Note:

Take-Home Experiment

This interesting activity examines the effect of weight upon terminal velocity. Gather together some nested coffee filters. Leaving them in their original shape, measure the time it takes for one, two, three, four, and five nested filters to fall to the floor from the same height (roughly 2 m). (Note that, due to the way the filters are nested, drag is constant and only mass varies.) They obtain terminal velocity quite quickly, so find this velocity as a function of mass. Plot the terminal velocity v versus mass. Also plot v^2 versus mass. Which of these relationships is more linear? What can you conclude from these graphs?

Example:

A Terminal Velocity

Find the terminal velocity of an 85-kg skydiver falling in a spread-eagle position.

Strategy

At terminal velocity, $F_{\text{net}} = 0$. Thus the drag force on the skydiver must equal the force of gravity (the person's weight). Using the equation of drag force, we find $mg = \frac{1}{2}\rho C A v^2$.

Thus the terminal velocity v_t can be written as

Equation:

$$v_t = \sqrt{\frac{2mg}{\rho C A}}.$$

Solution

All quantities are known except the person's projected area. This is an adult (85 kg) falling spread eagle. We can estimate the frontal area as

Equation:

$$A = (2 \text{ m})(0.35 \text{ m}) = 0.70 \text{ m}^2.$$

Using our equation for v_t , we find that

Equation:

$$\begin{aligned} v_t &= \sqrt{\frac{2(85 \text{ kg})(9.80 \text{ m/s}^2)}{(1.21 \text{ kg/m}^3)(1.0)(0.70 \text{ m}^2)}} \\ &= 44 \text{ m/s.} \end{aligned}$$

Discussion

This result is consistent with the value for v_t mentioned earlier. The 75-kg skydiver going feet first had a $v = 98 \text{ m/s}$. He weighed less but had a smaller frontal area and so a smaller drag due to the air.

The size of the object that is falling through air presents another interesting application of air drag. If you fall from a 5-m high branch of a tree, you will likely get hurt—possibly fracturing a bone. However, a small squirrel does this all the time, without getting hurt. You don't reach a terminal velocity in such a short distance, but the squirrel does.

The following interesting quote on animal size and terminal velocity is from a 1928 essay by a British biologist, J.B.S. Haldane, titled "On Being the Right Size."

To the mouse and any smaller animal, [gravity] presents practically no dangers. You can drop a mouse down a thousand-yard mine shaft; and, on arriving at the bottom, it gets a slight shock and walks away, provided that the ground is fairly soft. A rat is killed, a man is broken, and a horse splashes. For the resistance presented to movement by the air is proportional to the surface of the moving object. Divide an animal's length, breadth, and height each by ten; its weight is reduced to a thousandth, but its surface only to a hundredth. So the resistance to falling in the case of the small animal is relatively ten times greater than the driving force.

The above quadratic dependence of air drag upon velocity does not hold if the object is very small, is going very slow, or is in a denser medium than air. Then we find that the drag force is proportional just to the velocity. This relationship is given by **Stokes' law**, which states that

Equation:

$$F_s = 6\pi r\eta v,$$

where r is the radius of the object, η is the viscosity of the fluid, and v is the object's velocity.

Note:

Stokes' Law

Equation:

$$F_s = 6\pi r\eta v,$$

where r is the radius of the object, η is the viscosity of the fluid, and v is the object's velocity.

Good examples of this law are provided by microorganisms, pollen, and dust particles. Because each of these objects is so small, we find that many of these objects travel unaided only at a constant (terminal) velocity.

Terminal velocities for bacteria (size about $1\text{ }\mu\text{m}$) can be about $2\text{ }\mu\text{m/s}$. To

move at a greater speed, many bacteria swim using flagella (organelles shaped like little tails) that are powered by little motors embedded in the cell. Sediment in a lake can move at a greater terminal velocity (about $5 \mu\text{m/s}$), so it can take days to reach the bottom of the lake after being deposited on the surface.

If we compare animals living on land with those in water, you can see how drag has influenced evolution. Fishes, dolphins, and even massive whales are streamlined in shape to reduce drag forces. Birds are streamlined and migratory species that fly large distances often have particular features such as long necks. Flocks of birds fly in the shape of a spear head as the flock forms a streamlined pattern (see [\[link\]](#)). In humans, one important example of streamlining is the shape of sperm, which need to be efficient in their use of energy.



Geese fly in a V formation during their long migratory travels. This shape reduces drag and energy consumption for individual birds, and also allows them a better way to communicate.
(credit: Julo, Wikimedia Commons)

Note:**Galileo's Experiment**

Galileo is said to have dropped two objects of different masses from the Tower of Pisa. He measured how long it took each to reach the ground. Since stopwatches weren't readily available, how do you think he measured their fall time? If the objects were the same size, but with different masses, what do you think he should have observed? Would this result be different if done on the Moon?

Note:**PhET Explorations: Masses & Springs**

A realistic mass and spring laboratory. Hang masses from springs and adjust the spring stiffness and damping. You can even slow time. Transport the lab to different planets. A chart shows the kinetic, potential, and thermal energy for each spring.

[Masses](#)
[&](#)
[Spring](#)
[s](#)

Section Summary

- Drag forces acting on an object moving in a fluid oppose the motion. For larger objects (such as a baseball) moving at a velocity v in air, the drag force is given by

Equation:

$$F_D = \frac{1}{2} C \rho A v^2,$$

where C is the drag coefficient (typical values are given in [\[link\]](#)), A is the area of the object facing the fluid, and ρ is the fluid density.

- For small objects (such as a bacterium) moving in a denser medium (such as water), the drag force is given by Stokes' law,

Equation:

$$F_s = 6\pi\eta r v,$$

where r is the radius of the object, η is the fluid viscosity, and v is the object's velocity.

Conceptual Questions

Exercise:

Problem:

Athletes such as swimmers and bicyclists wear body suits in competition. Formulate a list of pros and cons of such suits.

Exercise:

Problem:

Two expressions were used for the drag force experienced by a moving object in a liquid. One depended upon the speed, while the other was proportional to the square of the speed. In which types of motion would each of these expressions be more applicable than the other one?

Exercise:

Problem:

As cars travel, oil and gasoline leaks onto the road surface. If a light rain falls, what does this do to the control of the car? Does a heavy rain make any difference?

Exercise:

Problem:

Why can a squirrel jump from a tree branch to the ground and run away undamaged, while a human could break a bone in such a fall?

Problems & Exercise**Exercise:****Problem:**

The terminal velocity of a person falling in air depends upon the weight and the area of the person facing the fluid. Find the terminal velocity (in meters per second and kilometers per hour) of an 80.0-kg skydiver falling in a pike (headfirst) position with a surface area of 0.140 m^2 .

Solution:

115 m/s; 414 km/hr

Exercise:**Problem:**

A 60-kg and a 90-kg skydiver jump from an airplane at an altitude of 6000 m, both falling in the pike position. Make some assumption on their frontal areas and calculate their terminal velocities. How long will it take for each skydiver to reach the ground (assuming the time to reach terminal velocity is small)? Assume all values are accurate to three significant digits.

Exercise:

Problem:

A 560-g squirrel with a surface area of 930 cm^2 falls from a 5.0-m tree to the ground. Estimate its terminal velocity. (Use a drag coefficient for a horizontal skydiver.) What will be the velocity of a 56-kg person hitting the ground, assuming no drag contribution in such a short distance?

Solution:

25 m/s; 9.9 m/s

Exercise:**Problem:**

To maintain a constant speed, the force provided by a car's engine must equal the drag force plus the force of friction of the road (the rolling resistance). (a) What are the magnitudes of drag forces at 70 km/h and 100 km/h for a Toyota Camry? (Drag area is 0.70 m^2) (b) What is the magnitude of drag force at 70 km/h and 100 km/h for a Hummer H2? (Drag area is 2.44 m^2) Assume all values are accurate to three significant digits.

Exercise:**Problem:**

By what factor does the drag force on a car increase as it goes from 65 to 110 km/h?

Solution:

2.9

Exercise:

Problem:

Calculate the speed a spherical rain drop would achieve falling from 5.00 km (a) in the absence of air drag (b) with air drag. Take the size across of the drop to be 4 mm, the density to be $1.00 \times 10^3 \text{ kg/m}^3$, and the surface area to be πr^2 .

Exercise:**Problem:**

Using Stokes' law, verify that the units for viscosity are kilograms per meter per second.

Solution:**Equation:**

$$[\eta] = \frac{[F_s]}{[r][v]} = \frac{\text{kg} \cdot \text{m/s}^2}{\text{m} \cdot \text{m/s}} = \frac{\text{kg}}{\text{m} \cdot \text{s}}$$

Exercise:**Problem:**

Find the terminal velocity of a spherical bacterium (diameter $2.00 \text{ }\mu\text{m}$) falling in water. You will first need to note that the drag force is equal to the weight at terminal velocity. Take the density of the bacterium to be $1.10 \times 10^3 \text{ kg/m}^3$.

Exercise:

Problem:

Stokes' law describes sedimentation of particles in liquids and can be used to measure viscosity. Particles in liquids achieve terminal velocity quickly. One can measure the time it takes for a particle to fall a certain distance and then use Stokes' law to calculate the viscosity of the liquid. Suppose a steel ball bearing (density $7.8 \times 10^3 \text{ kg/m}^3$, diameter 3.0 mm) is dropped in a container of motor oil. It takes 12 s to fall a distance of 0.60 m. Calculate the viscosity of the oil.

Solution:

0.76 kg/m · s

Glossary

drag force

F_D , found to be proportional to the square of the speed of the object; mathematically

Equation:

$$F_D \propto v^2$$

Equation:

$$F_D = \frac{1}{2} C \rho A v^2,$$

where C is the drag coefficient, A is the area of the object facing the fluid, and ρ is the density of the fluid

Stokes' law

$F_s = 6\pi r \eta v$, where r is the radius of the object, η is the viscosity of the fluid, and v is the object's velocity

Newton's Universal Law of Gravitation

- Explain Earth's gravitational force.
- Describe the gravitational effect of the Moon on Earth.
- Discuss weightlessness in space.
- Examine the Cavendish experiment

What do aching feet, a falling apple, and the orbit of the Moon have in common? Each is caused by the gravitational force. Our feet are strained by supporting our weight—the force of Earth's gravity on us. An apple falls from a tree because of the same force acting a few meters above Earth's surface. And the Moon orbits Earth because gravity is able to supply the necessary centripetal force at a distance of hundreds of millions of meters. In fact, the same force causes planets to orbit the Sun, stars to orbit the center of the galaxy, and galaxies to cluster together. Gravity is another example of underlying simplicity in nature. It is the weakest of the four basic forces found in nature, and in some ways the least understood. It is a force that acts at a distance, without physical contact, and is expressed by a formula that is valid everywhere in the universe, for masses and distances that vary from the tiny to the immense.

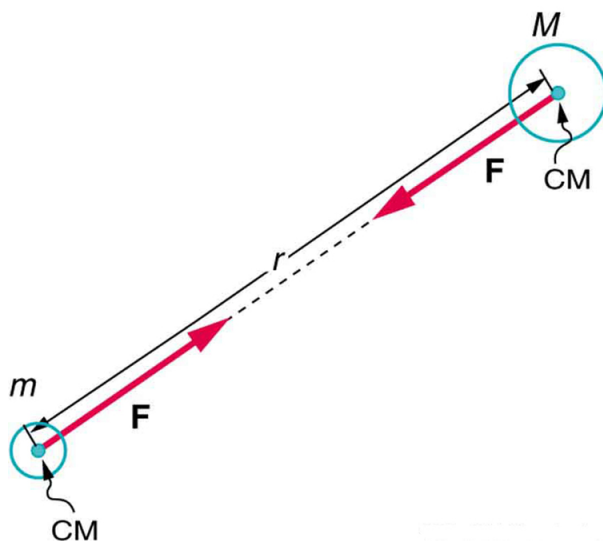
Sir Isaac Newton was the first scientist to precisely define the gravitational force, and to show that it could explain both falling bodies and astronomical motions. See [\[link\]](#). But Newton was not the first to suspect that the same force caused both our weight and the motion of planets. His forerunner Galileo Galilei had contended that falling bodies and planetary motions had the same cause. Some of Newton's contemporaries, such as Robert Hooke, Christopher Wren, and Edmund Halley, had also made some progress toward understanding gravitation. But Newton was the first to propose an exact mathematical form and to use that form to show that the motion of heavenly bodies should be conic sections—circles, ellipses, parabolas, and hyperbolas. This theoretical prediction was a major triumph—it had been known for some time that moons, planets, and comets follow such paths, but no one had been able to propose a mechanism that caused them to follow these paths and not others.



According to early accounts, Newton was inspired to make the connection between falling bodies and astronomical motions when he saw an apple fall from a tree and realized that if the gravitational force could extend above the ground to a tree, it might also reach the Sun. The inspiration of Newton's apple is a part of worldwide folklore and may even be based in fact. Great importance is attached to it because Newton's

universal law of
gravitation and his
laws of motion
answered very old
questions about nature
and gave tremendous
support to the notion
of underlying
simplicity and unity in
nature. Scientists still
expect underlying
simplicity to emerge
from their ongoing
inquiries into nature.

The gravitational force is relatively simple. It is always attractive, and it depends only on the masses involved and the distance between them. Stated in modern language, **Newton's universal law of gravitation** states that every particle in the universe attracts every other particle with a force along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.



Gravitational attraction is along a line joining the centers of mass of these two bodies. The magnitude of the force is the same on each, consistent with Newton's third law.

Note:

Misconception Alert

The magnitude of the force on each object (one has larger mass than the other) is the same, consistent with Newton's third law.

The bodies we are dealing with tend to be large. To simplify the situation we assume that the body acts as if its entire mass is concentrated at one specific point called the **center of mass** (CM), which will be further explored in [Linear Momentum and Collisions](#). For two bodies having masses m and M with a distance r between their centers of mass, the equation for Newton's universal law of gravitation is

Equation:

$$F = G \frac{mM}{r^2},$$

where F is the magnitude of the gravitational force and G is a proportionality factor called the **gravitational constant**. G is a universal gravitational constant—that is, it is thought to be the same everywhere in the universe. It has been measured experimentally to be

Equation:

$$G = 6.674 \times 10^{-11} \frac{\text{N} \cdot \text{m}^2}{\text{kg}^2}$$

in SI units. Note that the units of G are such that a force in newtons is obtained from $F = G \frac{mM}{r^2}$, when considering masses in kilograms and distance in meters. For example, two 1.000 kg masses separated by 1.000 m will experience a gravitational attraction of 6.674×10^{-11} N. This is an extraordinarily small force. The small magnitude of the gravitational force is consistent with everyday experience. We are unaware that even large objects like mountains exert gravitational forces on us. In fact, our body weight is the force of attraction of the *entire Earth* on us with a mass of 6×10^{24} kg.

Recall that the acceleration due to gravity g is about 9.80 m/s^2 on Earth. We can now determine why this is so. The weight of an object mg is the gravitational force between it and Earth. Substituting mg for F in Newton's universal law of gravitation gives

Equation:

$$mg = G \frac{mM}{r^2},$$

where m is the mass of the object, M is the mass of Earth, and r is the distance to the center of Earth (the distance between the centers of mass of the object and Earth). See [\[link\]](#). The mass m of the object cancels, leaving an equation for g :

Equation:

$$g = G \frac{M}{r^2}.$$

Substituting known values for Earth's mass and radius (to three significant figures),

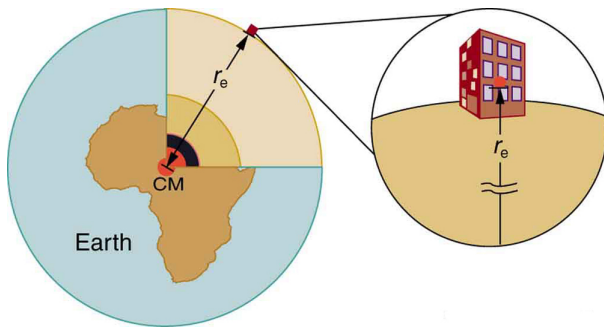
Equation:

$$g = \left(6.67 \times 10^{-11} \frac{\text{N} \cdot \text{m}^2}{\text{kg}^2} \right) \times \frac{5.98 \times 10^{24} \text{ kg}}{(6.38 \times 10^6 \text{ m})^2},$$

and we obtain a value for the acceleration of a falling body:

Equation:

$$g = 9.80 \text{ m/s}^2.$$



The distance between the centers of mass of Earth and an object on its surface is very nearly the same as the radius of Earth, because Earth is so much larger than the object.

This is the expected value *and is independent of the body's mass*. Newton's law of gravitation takes Galileo's observation that all masses fall with the same acceleration a step further, explaining the observation in terms of a force that causes objects to fall—in fact, in terms of a universally existing force of attraction between masses.

Note:

Take-Home Experiment

Take a marble, a ball, and a spoon and drop them from the same height. Do they hit the floor at the same time? If you drop a piece of paper as well, does it behave like the other objects? Explain your observations.

Note:**Making Connections**

Attempts are still being made to understand the gravitational force. As we shall see in [Particle Physics](#), modern physics is exploring the connections of gravity to other forces, space, and time. General relativity alters our view of gravitation, leading us to think of gravitation as bending space and time.

In the following example, we make a comparison similar to one made by Newton himself. He noted that if the gravitational force caused the Moon to orbit Earth, then the acceleration due to gravity should equal the centripetal acceleration of the Moon in its orbit. Newton found that the two accelerations agreed “pretty nearly.”

Example:**Earth’s Gravitational Force Is the Centripetal Force Making the Moon Move in a Curved Path**

- (a) Find the acceleration due to Earth’s gravity at the distance of the Moon.
- (b) Calculate the centripetal acceleration needed to keep the Moon in its orbit (assuming a circular orbit about a fixed Earth), and compare it with the value of the acceleration due to Earth’s gravity that you have just found.

Strategy for (a)

This calculation is the same as the one finding the acceleration due to gravity at Earth’s surface, except that r is the distance from the center of Earth to the center of the Moon. The radius of the Moon’s nearly circular orbit is 3.84×10^8 m.

Solution for (a)

Substituting known values into the expression for g found above, remembering that M is the mass of Earth not the Moon, yields

Equation:

$$\begin{aligned}
 g &= G \frac{M}{r^2} = \left(6.67 \times 10^{-11} \frac{\text{N} \cdot \text{m}^2}{\text{kg}^2} \right) \times \frac{5.98 \times 10^{24} \text{ kg}}{(3.84 \times 10^8 \text{ m})^2} \\
 &= 2.70 \times 10^{-3} \text{ m/s}^2
 \end{aligned}$$

Strategy for (b)

Centripetal acceleration can be calculated using either form of

Equation:

$$\left. \begin{aligned} a_c &= \frac{v^2}{r} \\ a_c &= r\omega^2 \end{aligned} \right\}$$

We choose to use the second form:

Equation:

$$a_c = r\omega^2,$$

where ω is the angular velocity of the Moon about Earth.

Solution for (b)

Given that the period (the time it takes to make one complete rotation) of the Moon's orbit is 27.3 days, (d) and using

Equation:

$$1 \text{ d} \times 24 \frac{\text{hr}}{\text{d}} \times 60 \frac{\text{min}}{\text{hr}} \times 60 \frac{\text{s}}{\text{min}} = 86,400 \text{ s}$$

we see that

Equation:

$$\omega = \frac{\Delta\theta}{\Delta t} = \frac{2\pi \text{ rad}}{(27.3 \text{ d})(86,400 \text{ s/d})} = 2.66 \times 10^{-6} \frac{\text{rad}}{\text{s}}.$$

The centripetal acceleration is

Equation:

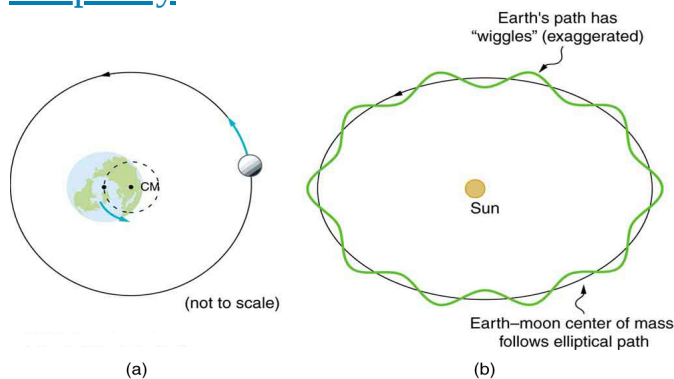
$$\begin{aligned}
 a_c &= r\omega^2 = (3.84 \times 10^8 \text{ m})(2.66 \times 10^{-6} \text{ rad/s})^2 \\
 &= 2.72 \times 10^{-3} \text{ m/s}^2
 \end{aligned}$$

The direction of the acceleration is toward the center of the Earth.

Discussion

The centripetal acceleration of the Moon found in (b) differs by less than 1% from the acceleration due to Earth's gravity found in (a). This agreement is approximate because the Moon's orbit is slightly elliptical, and Earth is not stationary (rather the Earth-Moon system rotates about its center of mass, which is located some 1700 km below Earth's surface). The clear implication is that Earth's gravitational force causes the Moon to orbit Earth.

Why does Earth not remain stationary as the Moon orbits it? This is because, as expected from Newton's third law, if Earth exerts a force on the Moon, then the Moon should exert an equal and opposite force on Earth (see [\[link\]](#)). We do not sense the Moon's effect on Earth's motion, because the Moon's gravity moves our bodies right along with Earth but there are other signs on Earth that clearly show the effect of the Moon's gravitational force as discussed in [Satellites and Kepler's Laws: An Argument for Simplicity](#).

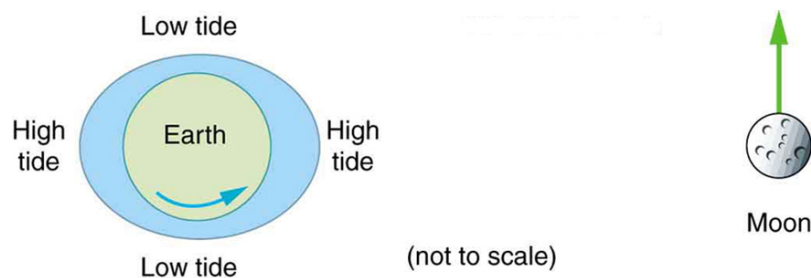


(a) Earth and the Moon rotate approximately once a month around their common center of mass. (b) Their center of mass orbits the Sun in an elliptical orbit, but Earth's path around the Sun has "wiggles" in it. Similar wiggles in the paths of stars have been observed and are

considered direct evidence of planets orbiting those stars. This is important because the planets' reflected light is often too dim to be observed.

Tides

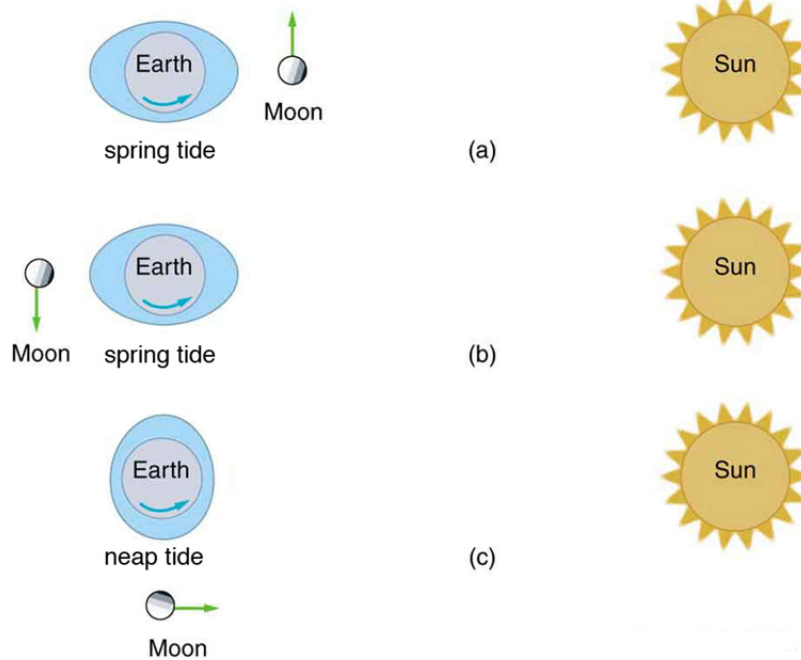
Ocean tides are one very observable result of the Moon's gravity acting on Earth. [\[link\]](#) is a simplified drawing of the Moon's position relative to the tides. Because water easily flows on Earth's surface, a high tide is created on the side of Earth nearest to the Moon, where the Moon's gravitational pull is strongest. Why is there also a high tide on the opposite side of Earth? The answer is that Earth is pulled toward the Moon more than the water on the far side, because Earth is closer to the Moon. So the water on the side of Earth closest to the Moon is pulled away from Earth, and Earth is pulled away from water on the far side. As Earth rotates, the tidal bulge (an effect of the tidal forces between an orbiting natural satellite and the primary planet that it orbits) keeps its orientation with the Moon. Thus there are two tides per day (the actual tidal period is about 12 hours and 25.2 minutes), because the Moon moves in its orbit each day as well).



The Moon causes ocean tides by attracting the water on the near side more than Earth, and by attracting Earth more than the water on the far side. The distances and sizes are

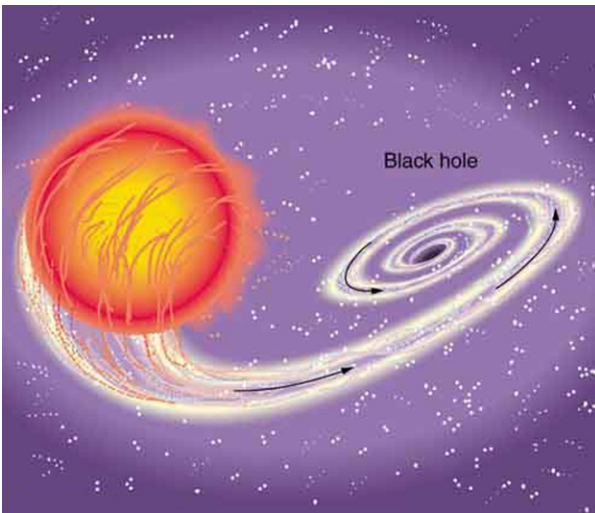
not to scale. For this simplified representation of the Earth-Moon system, there are two high and two low tides per day at any location, because Earth rotates under the tidal bulge.

The Sun also affects tides, although it has about half the effect of the Moon. However, the largest tides, called spring tides, occur when Earth, the Moon, and the Sun are aligned. The smallest tides, called neap tides, occur when the Sun is at a 90° angle to the Earth-Moon alignment.



(a, b) Spring tides: The highest tides occur when Earth, the Moon, and the Sun are aligned. (c) Neap tide: The lowest tides occur when the Sun lies at 90° to the Earth-Moon alignment. Note that this figure is not drawn to scale.

Tides are not unique to Earth but occur in many astronomical systems. The most extreme tides occur where the gravitational force is the strongest and varies most rapidly, such as near black holes (see [\[link\]](#)). A few likely candidates for black holes have been observed in our galaxy. These have masses greater than the Sun but have diameters only a few kilometers across. The tidal forces near them are so great that they can actually tear matter from a companion star.



A black hole is an object with such strong gravity that not even light can escape it. This black hole was created by the supernova of one star in a two-star system. The tidal forces created by the black hole are so great that it tears matter from the companion star. This matter is compressed and heated as it is sucked into the black hole, creating light and X-rays observable from Earth.

”Weightlessness” and Microgravity

In contrast to the tremendous gravitational force near black holes is the apparent gravitational field experienced by astronauts orbiting Earth. What is the effect of “weightlessness” upon an astronaut who is in orbit for months? Or what about the effect of weightlessness upon plant growth? Weightlessness doesn’t mean that an astronaut is not being acted upon by the gravitational force. There is no “zero gravity” in an astronaut’s orbit. The term just means that the astronaut is in free-fall, accelerating with the acceleration due to gravity. If an elevator cable breaks, the passengers inside will be in free fall and will experience weightlessness. You can experience short periods of weightlessness in some rides in amusement parks.



Astronauts experiencing weightlessness on board the International Space Station. (credit: NASA)

Microgravity refers to an environment in which the apparent net acceleration of a body is small compared with that produced by Earth at its surface. Many interesting biology and physics topics have been studied over the past three decades in the presence of microgravity. Of immediate concern is the effect on astronauts of extended times in outer space, such as at the International Space Station. Researchers have observed that muscles will atrophy (waste away) in this environment. There is also a corresponding loss of bone mass. Study continues on cardiovascular adaptation to space flight. On Earth, blood pressure is usually higher in the feet than in the head, because the higher column of blood exerts a downward force on it, due to gravity. When standing, 70% of your blood is below the level of the heart, while in a horizontal position, just the opposite occurs. What difference does the absence of this pressure differential have upon the heart?

Some findings in human physiology in space can be clinically important to the management of diseases back on Earth. On a somewhat negative note, spaceflight is known to affect the human immune system, possibly making the crew members more vulnerable to infectious diseases. Experiments flown in space also have shown that some bacteria grow faster in microgravity than they do on Earth. However, on a positive note, studies indicate that microbial antibiotic production can increase by a factor of two in space-grown cultures. One hopes to be able to understand these mechanisms so that similar successes can be achieved on the ground. In another area of physics space research, inorganic crystals and protein crystals have been grown in outer space that have much higher quality than any grown on Earth, so crystallography studies on their structure can yield much better results.

Plants have evolved with the stimulus of gravity and with gravity sensors. Roots grow downward and shoots grow upward. Plants might be able to provide a life support system for long duration space missions by regenerating the atmosphere, purifying water, and producing food. Some studies have indicated that plant growth and development are not affected by gravity, but there is still uncertainty about structural changes in plants grown in a microgravity environment.

The Cavendish Experiment: Then and Now

As previously noted, the universal gravitational constant G is determined experimentally. This definition was first done accurately by Henry Cavendish (1731–1810), an English scientist, in 1798, more than 100 years after Newton published his universal law of gravitation. The measurement of G is very basic and important because it determines the strength of one of the four forces in nature. Cavendish's experiment was very difficult because he measured the tiny gravitational attraction between two ordinary-sized masses (tens of kilograms at most), using apparatus like that in [\[link\]](#). Remarkably, his value for G differs by less than 1% from the best modern value.

One important consequence of knowing G was that an accurate value for Earth's mass could finally be obtained. This was done by measuring the acceleration due to gravity as accurately as possible and then calculating the mass of Earth M from the relationship Newton's universal law of gravitation gives

Equation:

$$mg = G \frac{mM}{r^2},$$

where m is the mass of the object, M is the mass of Earth, and r is the distance to the center of Earth (the distance between the centers of mass of the object and Earth). See [\[link\]](#). The mass m of the object cancels, leaving an equation for g :

Equation:

$$g = G \frac{M}{r^2}.$$

Rearranging to solve for M yields

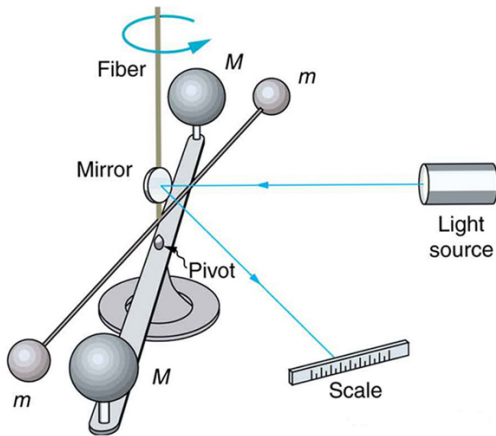
Equation:

$$M = \frac{gr^2}{G}.$$

So M can be calculated because all quantities on the right, including the radius of Earth r , are known from direct measurements. We shall see in [Satellites and Kepler's Laws: An Argument for Simplicity](#) that knowing G also allows for the determination of astronomical masses. Interestingly, of all the fundamental constants in physics, G is by far the least well determined.

The Cavendish experiment is also used to explore other aspects of gravity. One of the most interesting questions is whether the gravitational force depends on substance as well as mass—for example, whether one kilogram of lead exerts the same gravitational pull as one kilogram of water. A Hungarian scientist named Roland von Eötvös pioneered this inquiry early in the 20th century. He found, with an accuracy of five parts per billion, that the gravitational force does not depend on the substance. Such experiments continue today, and have improved upon Eötvös' measurements.

Cavendish-type experiments such as those of Eric Adelberger and others at the University of Washington, have also put severe limits on the possibility of a fifth force and have verified a major prediction of general relativity—that gravitational energy contributes to rest mass. Ongoing measurements there use a torsion balance and a parallel plate (not spheres, as Cavendish used) to examine how Newton's law of gravitation works over sub-millimeter distances. On this small-scale, do gravitational effects depart from the inverse square law? So far, no deviation has been observed.



Cavendish used an apparatus like this to measure the gravitational attraction between the two suspended spheres (m) and the two on the stand (M) by observing the amount of torsion (twisting) created in the fiber. Distance between the masses can be varied to check the dependence of the force on distance. Modern experiments of this type continue to explore gravity.

Section Summary

- Newton's universal law of gravitation: Every particle in the universe attracts every other particle with a force along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. In equation form, this is

Equation:

$$F = G \frac{mM}{r^2},$$

where F is the magnitude of the gravitational force. G is the gravitational constant, given by $G = 6.674 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$.

- Newton's law of gravitation applies universally.

Conceptual Questions**Exercise:****Problem:**

Action at a distance, such as is the case for gravity, was once thought to be illogical and therefore untrue. What is the ultimate determinant of the truth in physics, and why was this action ultimately accepted?

Exercise:**Problem:**

Two friends are having a conversation. Anna says a satellite in orbit is in freefall because the satellite keeps falling toward Earth. Tom says a satellite in orbit is not in freefall because the acceleration due to gravity is not 9.80 m/s^2 . Who do you agree with and why?

Exercise:**Problem:**

Draw a free body diagram for a satellite in an elliptical orbit showing why its speed increases as it approaches its parent body and decreases as it moves away.

Exercise:

Problem:

Newton's laws of motion and gravity were among the first to convincingly demonstrate the underlying simplicity and unity in nature. Many other examples have since been discovered, and we now expect to find such underlying order in complex situations. Is there proof that such order will always be found in new explorations?

Problem Exercises**Exercise:****Problem:**

(a) Calculate Earth's mass given the acceleration due to gravity at the North Pole is 9.830 m/s^2 and the radius of the Earth is 6371 km from center to pole.

(b) Compare this with the accepted value of $5.979 \times 10^{24} \text{ kg}$.

Solution:

a) $5.979 \times 10^{24} \text{ kg}$

b) This is identical to the best value to three significant figures.

Exercise:**Problem:**

(a) Calculate the magnitude of the acceleration due to gravity on the surface of Earth due to the Moon.

(b) Calculate the magnitude of the acceleration due to gravity at Earth due to the Sun.

(c) Take the ratio of the Moon's acceleration to the Sun's and comment on why the tides are predominantly due to the Moon in spite of this

number.

Exercise:

Problem:

- (a) What is the acceleration due to gravity on the surface of the Moon?
 - (b) On the surface of Mars? The mass of Mars is 6.418×10^{23} kg and its radius is 3.38×10^6 m.
-

Solution:

- a) 1.62 m/s^2
- b) 3.75 m/s^2

Exercise:

Problem:

- (a) Calculate the acceleration due to gravity on the surface of the Sun.
- (b) By what factor would your weight increase if you could stand on the Sun? (Never mind that you cannot.)

Exercise:

Problem:

The Moon and Earth rotate about their common center of mass, which is located about 4700 km from the center of Earth. (This is 1690 km below the surface.)

- (a) Calculate the magnitude of the acceleration due to the Moon's gravity at that point.
- (b) Calculate the magnitude of the centripetal acceleration of the center of Earth as it rotates about that point once each lunar month (about 27.3 d) and compare it with the acceleration found in part (a).

Comment on whether or not they are equal and why they should or should not be.

Solution:

a) $3.42 \times 10^{-5} \text{ m/s}^2$

b) $3.34 \times 10^{-5} \text{ m/s}^2$

The values are nearly identical. One would expect the gravitational force to be the same as the centripetal force at the core of the system.

Exercise:

Problem: Solve part (b) of [\[link\]](#) using $a_c = v^2/r$.

Exercise:

Problem:

Astrology, that unlikely and vague pseudoscience, makes much of the position of the planets at the moment of one's birth. The only known force a planet exerts on Earth is gravitational.

(a) Calculate the magnitude of the gravitational force exerted on a 4.20 kg baby by a 100 kg father 0.200 m away at birth (he is assisting, so he is close to the child).

(b) Calculate the magnitude of the force on the baby due to Jupiter if it is at its closest distance to Earth, some $6.29 \times 10^{11} \text{ m}$ away. How does the force of Jupiter on the baby compare to the force of the father on the baby? Other objects in the room and the hospital building also exert similar gravitational forces. (Of course, there could be an unknown force acting, but scientists first need to be convinced that there is even an effect, much less that an unknown force causes it.)

Solution:

a) $7.01 \times 10^{-7} \text{ N}$

b) $1.35 \times 10^{-6} \text{ N}$, 0.521

Exercise:

Problem:

The existence of the dwarf planet Pluto was proposed based on irregularities in Neptune's orbit. Pluto was subsequently discovered near its predicted position. But it now appears that the discovery was fortuitous, because Pluto is small and the irregularities in Neptune's orbit were not well known. To illustrate that Pluto has a minor effect on the orbit of Neptune compared with the closest planet to Neptune:

(a) Calculate the acceleration due to gravity at Neptune due to Pluto when they are $4.50 \times 10^{12} \text{ m}$ apart, as they are at present. The mass of Pluto is $1.4 \times 10^{22} \text{ kg}$.

(b) Calculate the acceleration due to gravity at Neptune due to Uranus, presently about $2.50 \times 10^{12} \text{ m}$ apart, and compare it with that due to Pluto. The mass of Uranus is $8.62 \times 10^{25} \text{ kg}$.

Exercise:

Problem:

(a) The Sun orbits the Milky Way galaxy once each $2.60 \times 10^8 \text{ y}$, with a roughly circular orbit averaging 3.00×10^4 light years in radius. (A light year is the distance traveled by light in 1 y.) Calculate the centripetal acceleration of the Sun in its galactic orbit. Does your result support the contention that a nearly inertial frame of reference can be located at the Sun?

(b) Calculate the average speed of the Sun in its galactic orbit. Does the answer surprise you?

Solution:

a) $1.66 \times 10^{-10} \text{ m/s}^2$

b) $2.17 \times 10^5 \text{ m/s}$

Exercise:

Problem: Unreasonable Result

A mountain 10.0 km from a person exerts a gravitational force on him equal to 2.00% of his weight.

- (a) Calculate the mass of the mountain.
 - (b) Compare the mountain's mass with that of Earth.
 - (c) What is unreasonable about these results?
 - (d) Which premises are unreasonable or inconsistent? (Note that accurate gravitational measurements can easily detect the effect of nearby mountains and variations in local geology.)
-

Solution:

a) $2.937 \times 10^{17} \text{ kg}$

b) 4.91×10^{-8}

of the Earth's mass.

c) The mass of the mountain and its fraction of the Earth's mass are too great.

d) The gravitational force assumed to be exerted by the mountain is too great.

Glossary

gravitational constant, G

a proportionality factor used in the equation for Newton's universal law of gravitation; it is a universal constant—that is, it is thought to be the same everywhere in the universe

center of mass

the point where the entire mass of an object can be thought to be concentrated

microgravity

an environment in which the apparent net acceleration of a body is small compared with that produced by Earth at its surface

Newton's universal law of gravitation

every particle in the universe attracts every other particle with a force along a line joining them; the force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them

Satellites and Kepler's Laws: An Argument for Simplicity

- State Kepler's laws of planetary motion.
- Derive the third Kepler's law for circular orbits.
- Discuss the Ptolemaic model of the universe.

Examples of gravitational orbits abound. Hundreds of artificial satellites orbit Earth together with thousands of pieces of debris. The Moon's orbit about Earth has intrigued humans from time immemorial. The orbits of planets, asteroids, meteors, and comets about the Sun are no less interesting. If we look further, we see almost unimaginable numbers of stars, galaxies, and other celestial objects orbiting one another and interacting through gravity.

All these motions are governed by gravitational force, and it is possible to describe them to various degrees of precision. Precise descriptions of complex systems must be made with large computers. However, we can describe an important class of orbits without the use of computers, and we shall find it instructive to study them. These orbits have the following characteristics:

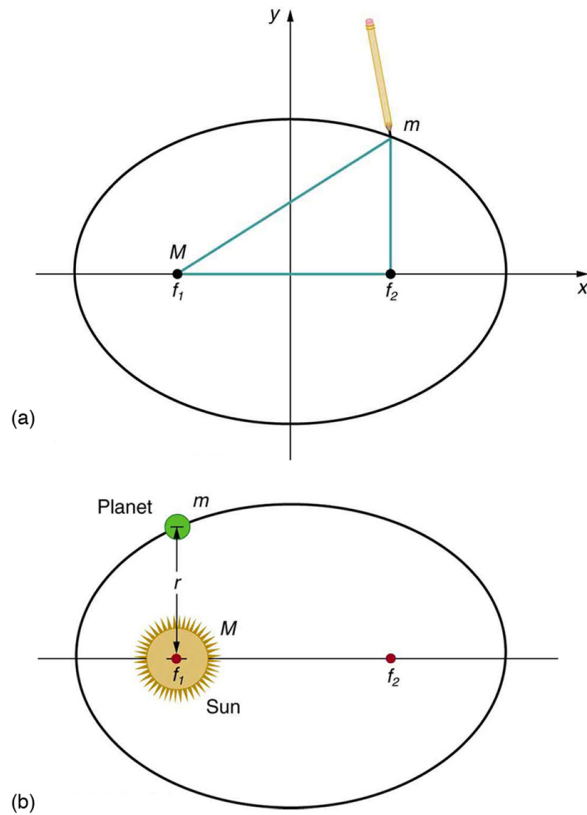
1. *A small mass m orbits a much larger mass M .* This allows us to view the motion as if M were stationary—in fact, as if from an inertial frame of reference placed on M —without significant error. Mass m is the satellite of M , if the orbit is gravitationally bound.
2. *The system is isolated from other masses.* This allows us to neglect any small effects due to outside masses.

The conditions are satisfied, to good approximation, by Earth's satellites (including the Moon), by objects orbiting the Sun, and by the satellites of other planets. Historically, planets were studied first, and there is a classical set of three laws, called Kepler's laws of planetary motion, that describe the orbits of all bodies satisfying the two previous conditions (not just planets in our solar system). These descriptive laws are named for the German astronomer Johannes Kepler (1571–1630), who devised them after careful study (over some 20 years) of a large amount of meticulously recorded observations of planetary motion done by Tycho Brahe (1546–1601). Such careful collection and detailed recording of methods and data are hallmarks of good science. Data constitute the evidence from which new interpretations and meanings can be constructed.

Kepler's Laws of Planetary Motion

Kepler's First Law

The orbit of each planet about the Sun is an ellipse with the Sun at one focus.



(a) An ellipse is a closed curve such that the sum of the distances from a point on the curve to the two foci (f_1 and f_2) is a constant. You can draw an ellipse as shown by putting a pin at each focus, and then placing a string around a pencil and the pins and tracing a line on paper. A circle is a special case of an ellipse in which the

two foci coincide (thus any point on the circle is the same distance from the center). (b)

For any closed gravitational orbit, m follows an elliptical path with M at one focus.

Kepler's first law states this fact for planets orbiting the Sun.

Kepler's Second Law

Each planet moves so that an imaginary line drawn from the Sun to the planet sweeps out equal areas in equal times (see [\[link\]](#)).

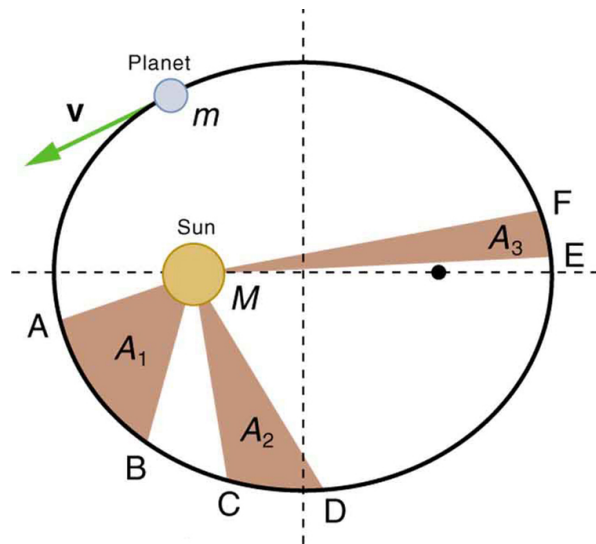
Kepler's Third Law

The ratio of the squares of the periods of any two planets about the Sun is equal to the ratio of the cubes of their average distances from the Sun. In equation form, this is

Equation:

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3},$$

where T is the period (time for one orbit) and r is the average radius. This equation is valid only for comparing two small masses orbiting the same large one. Most importantly, this is a descriptive equation only, giving no information as to the cause of the equality.



The shaded regions have equal areas. It takes equal times for m to go from A to B, from C to D, and from E to F. The mass m moves fastest when it is closest to M . Kepler's second law was originally devised for planets orbiting the Sun, but it has broader validity.

Note again that while, for historical reasons, Kepler's laws are stated for planets orbiting the Sun, they are actually valid for all bodies satisfying the two previously stated conditions.

Example:

Find the Time for One Orbit of an Earth Satellite

Given that the Moon orbits Earth each 27.3 d and that it is an average distance of 3.84×10^8 m from the center of Earth, calculate the period of an artificial satellite orbiting at an average altitude of 1500 km above Earth's surface.

Strategy

The period, or time for one orbit, is related to the radius of the orbit by Kepler's third law, given in mathematical form in $\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}$. Let us use the subscript 1 for the Moon and the subscript 2 for the satellite. We are asked to find T_2 . The given information tells us that the orbital radius of the Moon is $r_1 = 3.84 \times 10^8$ m, and that the period of the Moon is $T_1 = 27.3$ d. The height of the artificial satellite above Earth's surface is given, and so we must add the radius of Earth (6380 km) to get $r_2 = (1500 + 6380)$ km = 7880 km. Now all quantities are known, and so T_2 can be found.

Solution

Kepler's third law is

Equation:

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}.$$

To solve for T_2 , we cross-multiply and take the square root, yielding

Equation:

$$T_2^2 = T_1^2 \left(\frac{r_2}{r_1} \right)^3$$

Equation:

$$T_2 = T_1 \left(\frac{r_2}{r_1} \right)^{3/2}.$$

Substituting known values yields

Equation:

$$\begin{aligned} T_2 &= 27.3 \text{ d} \times \frac{24.0 \text{ h}}{\text{d}} \times \left(\frac{7880 \text{ km}}{3.84 \times 10^5 \text{ km}} \right)^{3/2} \\ &= 1.93 \text{ h.} \end{aligned}$$

Discussion This is a reasonable period for a satellite in a fairly low orbit. It is interesting that any satellite at this altitude will orbit in the same amount

of time. This fact is related to the condition that the satellite's mass is small compared with that of Earth.

People immediately search for deeper meaning when broadly applicable laws, like Kepler's, are discovered. It was Newton who took the next giant step when he proposed the law of universal gravitation. While Kepler was able to discover *what* was happening, Newton discovered that gravitational force was the cause.

Derivation of Kepler's Third Law for Circular Orbits

We shall derive Kepler's third law, starting with Newton's laws of motion and his universal law of gravitation. The point is to demonstrate that the force of gravity is the cause for Kepler's laws (although we will only derive the third one).

Let us consider a circular orbit of a small mass m around a large mass M , satisfying the two conditions stated at the beginning of this section. Gravity supplies the centripetal force to mass m . Starting with Newton's second law applied to circular motion,

Equation:

$$F_{\text{net}} = ma_c = m \frac{v^2}{r}.$$

The net external force on mass m is gravity, and so we substitute the force of gravity for F_{net} :

Equation:

$$G \frac{mM}{r^2} = m \frac{v^2}{r}.$$

The mass m cancels, yielding

Equation:

$$G \frac{M}{r} = v^2.$$

The fact that m cancels out is another aspect of the oft-noted fact that at a given location all masses fall with the same acceleration. Here we see that at a given orbital radius r , all masses orbit at the same speed. (This was implied by the result of the preceding worked example.) Now, to get at Kepler's third law, we must get the period T into the equation. By definition, period T is the time for one complete orbit. Now the average speed v is the circumference divided by the period—that is,

Equation:

$$v = \frac{2\pi r}{T}.$$

Substituting this into the previous equation gives

Equation:

$$G \frac{M}{r} = \frac{4\pi^2 r^2}{T^2}.$$

Solving for T^2 yields

Equation:

$$T^2 = \frac{4\pi^2}{GM} r^3.$$

Using subscripts 1 and 2 to denote two different satellites, and taking the ratio of the last equation for satellite 1 to satellite 2 yields

Equation:

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}.$$

This is Kepler's third law. Note that Kepler's third law is valid only for comparing satellites of the same parent body, because only then does the mass of the parent body M cancel.

Now consider what we get if we solve $T^2 = \frac{4\pi^2}{GM} r^3$ for the ratio r^3/T^2 . We obtain a relationship that can be used to determine the mass M of a parent body from the orbits of its satellites:

Equation:

$$\frac{r^3}{T^2} = \frac{G}{4\pi^2} M.$$

If r and T are known for a satellite, then the mass M of the parent can be calculated. This principle has been used extensively to find the masses of heavenly bodies that have satellites. Furthermore, the ratio r^3/T^2 should be a constant for all satellites of the same parent body (because $r^3/T^2 = GM/4\pi^2$). (See [\[link\]](#)).

It is clear from [\[link\]](#) that the ratio of r^3/T^2 is constant, at least to the third digit, for all listed satellites of the Sun, and for those of Jupiter. Small variations in that ratio have two causes—uncertainties in the r and T data, and perturbations of the orbits due to other bodies. Interestingly, those perturbations can be—and have been—used to predict the location of new planets and moons. This is another verification of Newton's universal law of gravitation.

Note:

Making Connections

Newton's universal law of gravitation is modified by Einstein's general theory of relativity, as we shall see in [Particle Physics](#). Newton's gravity is not seriously in error—it was and still is an extremely good approximation for most situations. Einstein's modification is most noticeable in extremely large gravitational fields, such as near black holes. However, general relativity also explains such phenomena as small but long-known deviations of the orbit of the planet Mercury from classical predictions.

The Case for Simplicity

The development of the universal law of gravitation by Newton played a pivotal role in the history of ideas. While it is beyond the scope of this text to cover that history in any detail, we note some important points. The definition of planet set in 2006 by the International Astronomical Union (IAU) states that in the solar system, a planet is a celestial body that:

- 1. is in orbit around the Sun,
- 2. has sufficient mass to assume hydrostatic equilibrium and
- 3. has cleared the neighborhood around its orbit.

A non-satellite body fulfilling only the first two of the above criteria is classified as “dwarf planet.”

In 2006, Pluto was demoted to a ‘dwarf planet’ after scientists revised their definition of what constitutes a “true” planet.

Parent	Satellite	Average orbital radius $r(\text{km})$	Period $T(\text{y})$	$r^3 / T^2 (\text{km}^3 / \text{y}^2)$
Earth	Moon	3.84×10^5	0.07481	1.01×10^{19}
Sun	Mercury	5.79×10^7	0.2409	3.34×10^{24}

Parent	Satellite	Average orbital radius $r(\text{km})$	Period $T(\text{y})$	$r^3 / T^2 (\text{km}^3 / \text{y}^2)$
	Venus	1.082×10^8	0.6150	3.35×10^{24}
	Earth	1.496×10^8	1.000	3.35×10^{24}
	Mars	2.279×10^8	1.881	3.35×10^{24}
	Jupiter	7.783×10^8	11.86	3.35×10^{24}
	Saturn	1.427×10^9	29.46	3.35×10^{24}
	Neptune	4.497×10^9	164.8	3.35×10^{24}
	Pluto	5.90×10^9	248.3	3.33×10^{24}
Jupiter	Io	4.22×10^5	0.00485 (1.77 d)	3.19×10^{21}
	Europa	6.71×10^5	0.00972 (3.55 d)	3.20×10^{21}

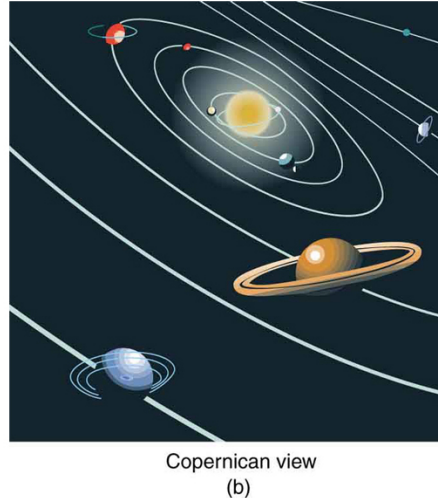
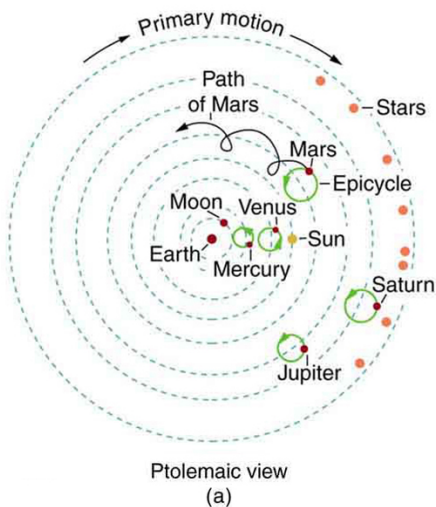
Parent	Satellite	Average orbital radius $r(\text{km})$	Period $T(\text{y})$	$r^3 / T^2 (\text{km}^3 / \text{y}^2)$
	Ganymede	1.07×10^6	0.0196 (7.16 d)	3.19×10^{21}
	Callisto	1.88×10^6	0.0457 (16.19 d)	3.20×10^{21}

Orbital Data and Kepler's Third Law

The universal law of gravitation is a good example of a physical principle that is very broadly applicable. That single equation for the gravitational force describes all situations in which gravity acts. It gives a cause for a vast number of effects, such as the orbits of the planets and moons in the solar system. It epitomizes the underlying unity and simplicity of physics.

Before the discoveries of Kepler, Copernicus, Galileo, Newton, and others, the solar system was thought to revolve around Earth as shown in [\[link\]](#)(a). This is called the Ptolemaic view, for the Greek philosopher who lived in the second century AD. This model is characterized by a list of facts for the motions of planets with no cause and effect explanation. There tended to be a different rule for each heavenly body and a general lack of simplicity.

[\[link\]](#)(b) represents the modern or Copernican model. In this model, a small set of rules and a single underlying force explain not only all motions in the solar system, but all other situations involving gravity. The breadth and simplicity of the laws of physics are compelling. As our knowledge of nature has grown, the basic simplicity of its laws has become ever more evident.



(a) The Ptolemaic model of the universe has Earth at the center with the Moon, the planets, the Sun, and the stars revolving about it in complex superpositions of circular paths. This geocentric model, which can be made progressively more accurate by adding more circles, is purely descriptive, containing no hints as to what are the causes of these motions. (b) The Copernican model has the Sun at the center of the solar system. It is fully explained by a small number of laws of physics, including Newton's universal law of gravitation.

Section Summary

- Kepler's laws are stated for a small mass m orbiting a larger mass M in near-isolation. Kepler's laws of planetary motion are then as follows:

Kepler's first law

The orbit of each planet about the Sun is an ellipse with the Sun at one focus.

Kepler's second law

Each planet moves so that an imaginary line drawn from the Sun to the planet sweeps out equal areas in equal times.

Kepler's third law

The ratio of the squares of the periods of any two planets about the Sun is equal to the ratio of the cubes of their average distances from the Sun:

Equation:

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3},$$

where T is the period (time for one orbit) and r is the average radius of the orbit.

- The period and radius of a satellite's orbit about a larger body M are related by

Equation:

$$T^2 = \frac{4\pi^2}{GM} r^3$$

or

Equation:

$$\frac{r^3}{T^2} = \frac{G}{4\pi^2} M.$$

Conceptual Questions

Exercise:

Problem:

In what frame(s) of reference are Kepler's laws valid? Are Kepler's laws purely descriptive, or do they contain causal information?

Problem Exercises

Exercise:

Problem:

A geosynchronous Earth satellite is one that has an orbital period of precisely 1 day. Such orbits are useful for communication and weather observation because the satellite remains above the same point on Earth (provided it orbits in the equatorial plane in the same direction as Earth's rotation). Calculate the radius of such an orbit based on the data for the moon in [\[link\]](#).

Exercise:

Problem:

Calculate the mass of the Sun based on data for Earth's orbit and compare the value obtained with the Sun's actual mass.

Solution:

$$1.98 \times 10^{30} \text{ kg}$$

Exercise:

Problem:

Find the mass of Jupiter based on data for the orbit of one of its moons, and compare your result with its actual mass.

Exercise:

Problem:

Find the ratio of the mass of Jupiter to that of Earth based on data in [\[link\]](#).

Solution:

$$\frac{M_J}{M_E} = 316$$

Exercise:

Problem:

Astronomical observations of our Milky Way galaxy indicate that it has a mass of about 8.0×10^{11} solar masses. A star orbiting on the galaxy's periphery is about 6.0×10^4 light years from its center. (a) What should the orbital period of that star be? (b) If its period is 6.0×10^7 years instead, what is the mass of the galaxy? Such calculations are used to imply the existence of "dark matter" in the universe and have indicated, for example, the existence of very massive black holes at the centers of some galaxies.

Exercise:**Problem: Integrated Concepts**

Space debris left from old satellites and their launchers is becoming a hazard to other satellites. (a) Calculate the speed of a satellite in an orbit 900 km above Earth's surface. (b) Suppose a loose rivet is in an orbit of the same radius that intersects the satellite's orbit at an angle of 90° relative to Earth. What is the velocity of the rivet relative to the satellite just before striking it? (c) Given the rivet is 3.00 mm in size, how long will its collision with the satellite last? (d) If its mass is 0.500 g, what is the average force it exerts on the satellite? (e) How much energy in joules is generated by the collision? (The satellite's velocity does not change appreciably, because its mass is much greater than the rivet's.)

Solution:

a) $7.4 \times 10^3 \text{ m/s}$

b) $1.05 \times 10^3 \text{ m/s}$

c) $2.86 \times 10^{-7} \text{ s}$

d) $1.84 \times 10^7 \text{ N}$

e) $2.76 \times 10^4 \text{ J}$

Exercise:

Problem: Unreasonable Results

(a) Based on Kepler's laws and information on the orbital characteristics of the Moon, calculate the orbital radius for an Earth satellite having a period of 1.00 h. (b) What is unreasonable about this result? (c) What is unreasonable or inconsistent about the premise of a 1.00 h orbit?

Solution:

a) $5.08 \times 10^3 \text{ km}$

b) This radius is unreasonable because it is less than the radius of earth.

c) The premise of a one-hour orbit is inconsistent with the known radius of the earth.

Exercise:**Problem: Construct Your Own Problem**

On February 14, 2000, the NEAR spacecraft was successfully inserted into orbit around Eros, becoming the first artificial satellite of an asteroid. Construct a problem in which you determine the orbital speed for a satellite near Eros. You will need to find the mass of the asteroid and consider such things as a safe distance for the orbit. Although Eros is not spherical, calculate the acceleration due to gravity on its surface at a point an average distance from its center of mass. Your instructor may also wish to have you calculate the escape velocity from this point on Eros.

Introduction to Work, Energy, and Energy Resources

class="introduction"

How many
forms of
energy can
you identify
in this
photograph
of a wind
farm in
Iowa?
(credit:
Jürgen from
Sandesneben
, Germany,
Wikimedia
Commons)



Energy plays an essential role both in everyday events and in scientific phenomena. You can no doubt name many forms of energy, from that provided by our foods, to the energy we use to run our cars, to the sunlight that warms us on the beach. You can also cite examples of what people call energy that may not be scientific, such as someone having an energetic personality. Not only does energy have many interesting forms, it is

involved in almost all phenomena, and is one of the most important concepts of physics. What makes it even more important is that the total amount of energy in the universe is constant. Energy can change forms, but it cannot appear from nothing or disappear without a trace. Energy is thus one of a handful of physical quantities that we say is *conserved*.

Conservation of energy (as physicists like to call the principle that energy can neither be created nor destroyed) is based on experiment. Even as scientists discovered new forms of energy, conservation of energy has always been found to apply. Perhaps the most dramatic example of this was supplied by Einstein when he suggested that mass is equivalent to energy (his famous equation $E = mc^2$).

From a societal viewpoint, energy is one of the major building blocks of modern civilization. Energy resources are key limiting factors to economic growth. The world use of energy resources, especially oil, continues to grow, with ominous consequences economically, socially, politically, and environmentally. We will briefly examine the world's energy use patterns at the end of this chapter.

There is no simple, yet accurate, scientific definition for energy. Energy is characterized by its many forms and the fact that it is conserved. We can loosely define **energy** as the ability to do work, admitting that in some circumstances not all energy is available to do work. Because of the association of energy with work, we begin the chapter with a discussion of work. Work is intimately related to energy and how energy moves from one system to another or changes form.

Work: The Scientific Definition

- Explain how an object must be displaced for a force on it to do work.
- Explain how relative directions of force and displacement determine whether the work done is positive, negative, or zero.

What It Means to Do Work

The scientific definition of work differs in some ways from its everyday meaning. Certain things we think of as hard work, such as writing an exam or carrying a heavy load on level ground, are not work as defined by a scientist. The scientific definition of work reveals its relationship to energy—whenever work is done, energy is transferred.

For work, in the scientific sense, to be done, a force must be exerted and there must be displacement in the direction of the force.

Formally, the **work** done on a system by a constant force is defined to be *the product of the component of the force in the direction of motion times the distance through which the force acts*. For one-way motion in one dimension, this is expressed in equation form as

Equation:

$$W = | \mathbf{F} | (\cos \theta) | \mathbf{d} |,$$

where W is work, \mathbf{d} is the displacement of the system, and θ is the angle between the force vector \mathbf{F} and the displacement vector \mathbf{d} , as in [\[link\]](#). We can also write this as

Equation:

$$W = Fd \cos \theta.$$

To find the work done on a system that undergoes motion that is not one-way or that is in two or three dimensions, we divide the motion into one-way one-dimensional segments and add up the work done over each segment.

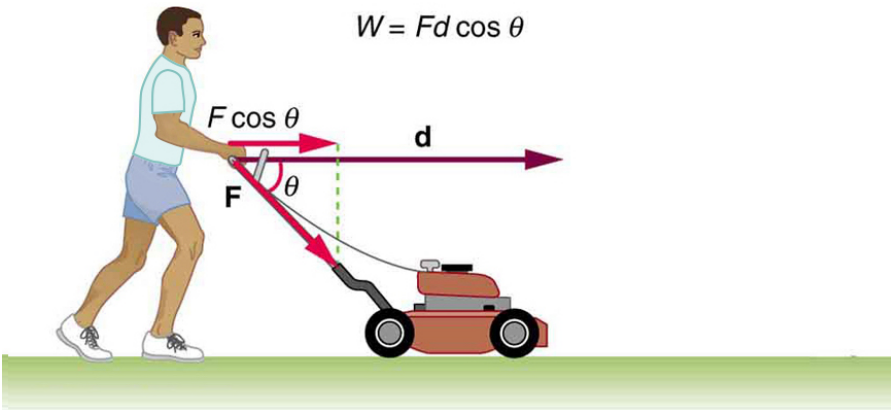
Note:**What is Work?**

The work done on a system by a constant force is *the product of the component of the force in the direction of motion times the distance through which the force acts*. For one-way motion in one dimension, this is expressed in equation form as

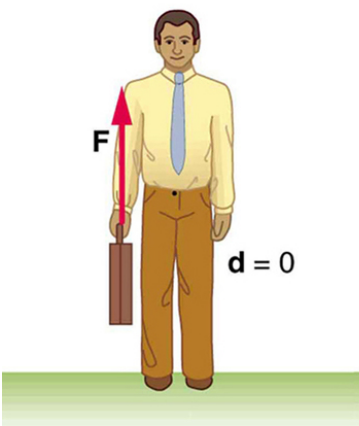
Equation:

$$W = Fd \cos \theta,$$

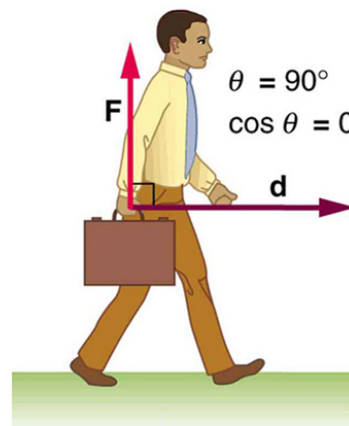
where W is work, F is the magnitude of the force on the system, d is the magnitude of the displacement of the system, and θ is the angle between the force vector \mathbf{F} and the displacement vector \mathbf{d} .



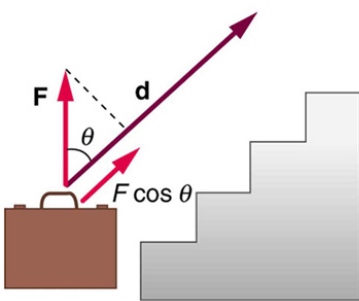
(a)



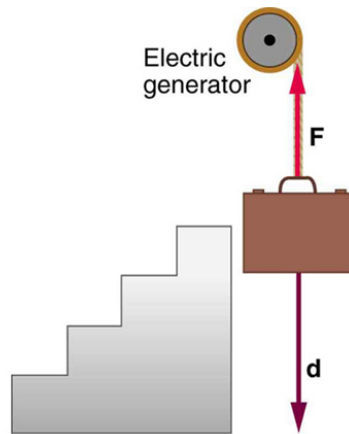
(b)



(c)



(d)



(e)

Examples of work. (a) The work done by the force \mathbf{F} on this lawn mower is $Fd \cos \theta$. Note that $F \cos \theta$ is the component of the force in the direction of motion. (b) A person holding a briefcase does no work on it, because there is no

displacement. No energy is transferred to or from the briefcase. (c) The person moving the briefcase horizontally at a constant speed does no work on it, and transfers no energy to it. (d) Work is done on the briefcase by carrying it up stairs at constant speed, because there is necessarily a component of force \mathbf{F} in the direction of the motion. Energy is transferred to the briefcase and could in turn be used to do work. (e) When the briefcase is lowered, energy is transferred out of the briefcase and into an electric generator. Here the work done on the briefcase by the generator is negative, removing energy from the briefcase, because \mathbf{F} and $d\mathbf{l}$ are in opposite directions.

To examine what the definition of work means, let us consider the other situations shown in [\[link\]](#). The person holding the briefcase in [\[link\]\(b\)](#) does no work, for example. Here $d = 0$, so $W = 0$. Why is it you get tired just holding a load? The answer is that your muscles are doing work against one another, *but they are doing no work on the system of interest* (the “briefcase-Earth system”—see [Gravitational Potential Energy](#) for more details). There must be displacement for work to be done, and there must be a component of the force in the direction of the motion. For example, the person carrying the briefcase on level ground in [\[link\]\(c\)](#) does no work on it, because the force is perpendicular to the motion. That is, $\cos 90^\circ = 0$, and so $W = 0$.

In contrast, when a force exerted on the system has a component in the direction of motion, such as in [\[link\]\(d\)](#), work is done—energy is transferred to the briefcase. Finally, in [\[link\]\(e\)](#), energy is transferred from the briefcase to a generator. There are two good ways to interpret this energy transfer. One interpretation is that the briefcase’s weight does work on the generator, giving it energy. The other interpretation is that the generator does negative work on the briefcase, thus removing energy from it. The drawing shows the latter, with the force from the generator upward

on the briefcase, and the displacement downward. This makes $\theta = 180^\circ$, and $\cos 180^\circ = -1$; therefore, W is negative.

Calculating Work

Work and energy have the same units. From the definition of work, we see that those units are force times distance. Thus, in SI units, work and energy are measured in **newton-meters**. A newton-meter is given the special name **joule** (J), and $1 \text{ J} = 1 \text{ N} \cdot \text{m} = 1 \text{ kg} \cdot \text{m}^2/\text{s}^2$. One joule is not a large amount of energy; it would lift a small 100-gram apple a distance of about 1 meter.

Example:

Calculating the Work You Do to Push a Lawn Mower Across a Large Lawn

How much work is done on the lawn mower by the person in [\[link\]](#)(a) if he exerts a constant force of 75.0 N at an angle 35° below the horizontal and pushes the mower 25.0 m on level ground? Convert the amount of work from joules to kilocalories and compare it with this person's average daily intake of 10,000 kJ (about 2400 kcal) of food energy. One *calorie* (1 cal) of heat is the amount required to warm 1 g of water by 1°C , and is equivalent to 4.184 J, while one *food calorie* (1 kcal) is equivalent to 4184 J.

Strategy

We can solve this problem by substituting the given values into the definition of work done on a system, stated in the equation $W = Fd \cos \theta$. The force, angle, and displacement are given, so that only the work W is unknown.

Solution

The equation for the work is

Equation:

$$W = Fd \cos \theta.$$

Substituting the known values gives

Equation:

$$\begin{aligned} W &= (75.0 \text{ N})(25.0 \text{ m}) \cos (35.0^\circ) \\ &= 1536 \text{ J} = 1.54 \times 10^3 \text{ J}. \end{aligned}$$

Converting the work in joules to kilocalories yields

$W = (1536 \text{ J})(1 \text{ kcal}/4184 \text{ J}) = 0.367 \text{ kcal}$. The ratio of the work done to the daily consumption is

Equation:

$$\frac{W}{2400 \text{ kcal}} = 1.53 \times 10^{-4}.$$

Discussion

This ratio is a tiny fraction of what the person consumes, but it is typical. Very little of the energy released in the consumption of food is used to do work. Even when we “work” all day long, less than 10% of our food energy intake is used to do work and more than 90% is converted to thermal energy or stored as chemical energy in fat.

Section Summary

- Work is the transfer of energy by a force acting on an object as it is displaced.
- The work W that a force \mathbf{F} does on an object is the product of the magnitude F of the force, times the magnitude d of the displacement, times the cosine of the angle θ between them. In symbols,

Equation:

$$W = Fd \cos \theta.$$

- The SI unit for work and energy is the joule (J), where $1 \text{ J} = 1 \text{ N} \cdot \text{m} = 1 \text{ kg} \cdot \text{m}^2/\text{s}^2$.
- The work done by a force is zero if the displacement is either zero or perpendicular to the force.

- The work done is positive if the force and displacement have the same direction, and negative if they have opposite direction.

Conceptual Questions

Exercise:

Problem:

Give an example of something we think of as work in everyday circumstances that is not work in the scientific sense. Is energy transferred or changed in form in your example? If so, explain how this is accomplished without doing work.

Exercise:

Problem:

Give an example of a situation in which there is a force and a displacement, but the force does no work. Explain why it does no work.

Exercise:

Problem:

Describe a situation in which a force is exerted for a long time but does no work. Explain.

Problems & Exercises

Exercise:

Problem:

How much work does a supermarket checkout attendant do on a can of soup he pushes 0.600 m horizontally with a force of 5.00 N? Express your answer in joules and kilocalories.

Solution:

Equation:

$$3.00 \text{ J} = 7.17 \times 10^{-4} \text{ kcal}$$

Exercise:**Problem:**

A 75.0-kg person climbs stairs, gaining 2.50 meters in height. Find the work done to accomplish this task.

Exercise:**Problem:**

(a) Calculate the work done on a 1500-kg elevator car by its cable to lift it 40.0 m at constant speed, assuming friction averages 100 N. (b) What is the work done on the lift by the gravitational force in this process? (c) What is the total work done on the lift?

Solution:

(a) $5.92 \times 10^5 \text{ J}$

(b) $-5.88 \times 10^5 \text{ J}$

(c) The net force is zero.

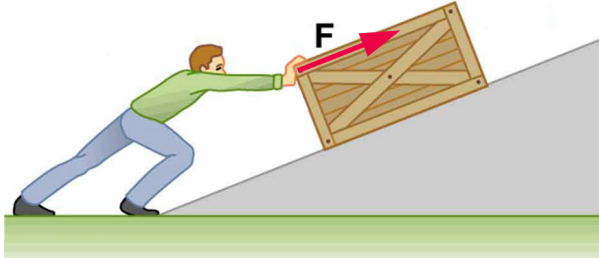
Exercise:**Problem:**

Suppose a car travels 108 km at a speed of 30.0 m/s, and uses 2.0 gal of gasoline. Only 30% of the gasoline goes into useful work by the force that keeps the car moving at constant speed despite friction. (See [\[link\]](#) for the energy content of gasoline.) (a) What is the magnitude of the force exerted to keep the car moving at constant speed? (b) If the required force is directly proportional to speed, how many gallons will be used to drive 108 km at a speed of 28.0 m/s?

Exercise:

Problem:

Calculate the work done by an 85.0-kg man who pushes a crate 4.00 m up along a ramp that makes an angle of 20.0° with the horizontal. (See [\[link\]](#).) He exerts a force of 500 N on the crate parallel to the ramp and moves at a constant speed. Be certain to include the work he does on the crate *and* on his body to get up the ramp.



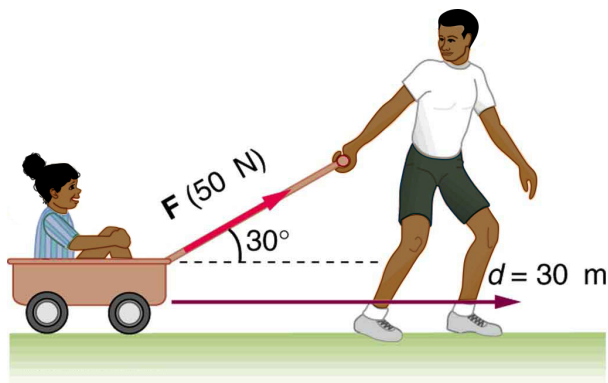
A man pushes a crate up a ramp.

Solution:**Equation:**

$$3.14 \times 10^3 \text{ J}$$

Exercise:**Problem:**

How much work is done by the boy pulling his sister 30.0 m in a wagon as shown in [\[link\]](#)? Assume no friction acts on the wagon.



The boy does work on the system of the wagon and the child when he pulls them as shown.

Exercise:

Problem:

A shopper pushes a grocery cart 20.0 m at constant speed on level ground, against a 35.0 N frictional force. He pushes in a direction 25.0° below the horizontal. (a) What is the work done on the cart by friction? (b) What is the work done on the cart by the gravitational force? (c) What is the work done on the cart by the shopper? (d) Find the force the shopper exerts, using energy considerations. (e) What is the total work done on the cart?

Solution:

(a) -700 J

(b) 0

(c) 700 J

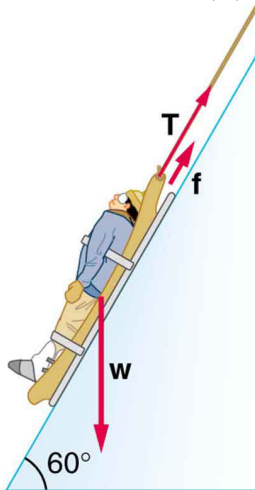
(d) 38.6 N

(e) 0

Exercise:

Problem:

Suppose the ski patrol lowers a rescue sled and victim, having a total mass of 90.0 kg , down a 60.0° slope at constant speed, as shown in [\[link\]](#). The coefficient of friction between the sled and the snow is 0.100 . (a) How much work is done by friction as the sled moves 30.0 m along the hill? (b) How much work is done by the rope on the sled in this distance? (c) What is the work done by the gravitational force on the sled? (d) What is the total work done?



A rescue sled and victim are lowered down a steep slope.

Glossary

energy

the ability to do work

work

the transfer of energy by a force that causes an object to be displaced;
the product of the component of the force in the direction of the
displacement and the magnitude of the displacement

joule

SI unit of work and energy, equal to one newton-meter

Kinetic Energy and the Work-Energy Theorem

- Explain work as a transfer of energy and net work as the work done by the net force.
- Explain and apply the work-energy theorem.

Work Transfers Energy

What happens to the work done on a system? Energy is transferred into the system, but in what form? Does it remain in the system or move on? The answers depend on the situation. For example, if the lawn mower in [\[link\]](#) (a) is pushed just hard enough to keep it going at a constant speed, then energy put into the mower by the person is removed continuously by friction, and eventually leaves the system in the form of heat transfer. In contrast, work done on the briefcase by the person carrying it up stairs in [\[link\]](#) (d) is stored in the briefcase-Earth system and can be recovered at any time, as shown in [\[link\]](#) (e). In fact, the building of the pyramids in ancient Egypt is an example of storing energy in a system by doing work on the system. Some of the energy imparted to the stone blocks in lifting them during construction of the pyramids remains in the stone-Earth system and has the potential to do work.

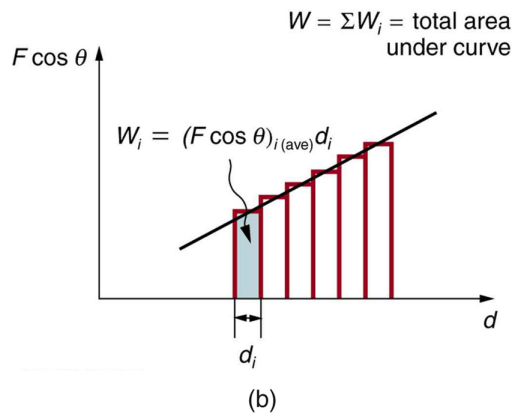
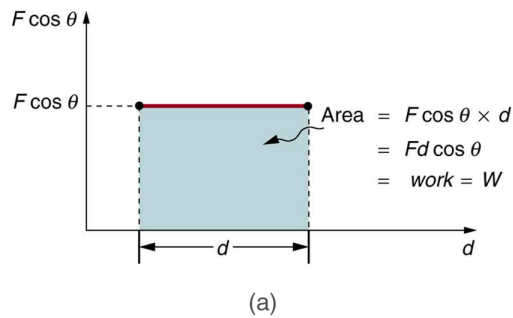
In this section we begin the study of various types of work and forms of energy. We will find that some types of work leave the energy of a system constant, for example, whereas others change the system in some way, such as making it move. We will also develop definitions of important forms of energy, such as the energy of motion.

Net Work and the Work-Energy Theorem

We know from the study of Newton's laws in [Dynamics: Force and Newton's Laws of Motion](#) that net force causes acceleration. We will see in this section that work done by the net force gives a system energy of motion, and in the process we will also find an expression for the energy of motion.

Let us start by considering the total, or net, work done on a system. Net work is defined to be the sum of work done by all external forces—that is, **net work** is the work done by the net external force \mathbf{F}_{net} . In equation form, this is $W_{\text{net}} = F_{\text{net}}d \cos \theta$ where θ is the angle between the force vector and the displacement vector.

[\[link\]](#)(a) shows a graph of force versus displacement for the component of the force in the direction of the displacement—that is, an $F \cos \theta$ vs. d graph. In this case, $F \cos \theta$ is constant. You can see that the area under the graph is $Fd \cos \theta$, or the work done. [\[link\]](#)(b) shows a more general process where the force varies. The area under the curve is divided into strips, each having an average force $(F \cos \theta)_{i(\text{ave})}$. The work done is $(F \cos \theta)_{i(\text{ave})}d_i$ for each strip, and the total work done is the sum of the W_i . Thus the total work done is the total area under the curve, a useful property to which we shall refer later.

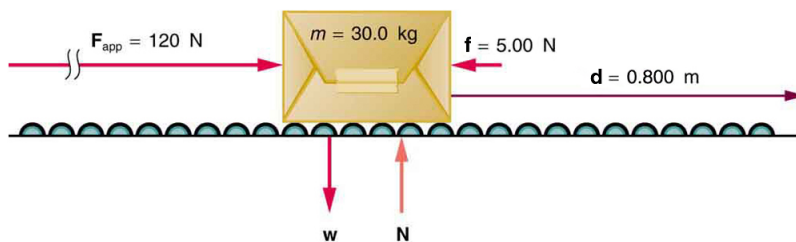


(a) A graph of $F \cos \theta$ vs. d , when $F \cos \theta$ is

constant. The area under the curve represents the work done by the force.

(b) A graph of $F \cos \theta$ vs. d in which the force varies. The work done for each interval is the area of each strip; thus, the total area under the curve equals the total work done.

Net work will be simpler to examine if we consider a one-dimensional situation where a force is used to accelerate an object in a direction parallel to its initial velocity. Such a situation occurs for the package on the roller belt conveyor system shown in [\[link\]](#).



A package on a roller belt is pushed horizontally through a distance d .

The force of gravity and the normal force acting on the package are perpendicular to the displacement and do no work. Moreover, they are also equal in magnitude and opposite in direction so they cancel in calculating the net force. The net force arises solely from the horizontal applied force F_{app} and the horizontal friction force f . Thus, as expected, the net force is

parallel to the displacement, so that $\theta = 0^\circ$ and $\cos \theta = 1$, and the net work is given by

Equation:

$$W_{\text{net}} = F_{\text{net}}d.$$

The effect of the net force \mathbf{F}_{net} is to accelerate the package from v_0 to v . The kinetic energy of the package increases, indicating that the net work done on the system is positive. (See [\[link\]](#).) By using Newton's second law, and doing some algebra, we can reach an interesting conclusion. Substituting $F_{\text{net}} = ma$ from Newton's second law gives

Equation:

$$W_{\text{net}} = mad.$$

To get a relationship between net work and the speed given to a system by the net force acting on it, we take $d = x - x_0$ and use the equation studied in [Motion Equations for Constant Acceleration in One Dimension](#) for the change in speed over a distance d if the acceleration has the constant value a ; namely, $v^2 = v_0^2 + 2ad$ (note that a appears in the expression for the net work). Solving for acceleration gives $a = \frac{v^2 - v_0^2}{2d}$. When a is substituted into the preceding expression for W_{net} , we obtain

Equation:

$$W_{\text{net}} = m \left(\frac{v^2 - v_0^2}{2d} \right) d.$$

The d cancels, and we rearrange this to obtain

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

This expression is called the **work-energy theorem**, and it actually applies *in general* (even for forces that vary in direction and magnitude), although we have derived it for the special case of a constant force parallel to the displacement. The theorem implies that the net work on a system equals the change in the quantity $\frac{1}{2}mv^2$. This quantity is our first example of a form of energy.

Note:

The Work-Energy Theorem

The net work on a system equals the change in the quantity $\frac{1}{2}mv^2$.

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$$

The quantity $\frac{1}{2}mv^2$ in the work-energy theorem is defined to be the translational **kinetic energy** (KE) of a mass m moving at a speed v . (*Translational* kinetic energy is distinct from *rotational* kinetic energy, which is considered later.) In equation form, the translational kinetic energy,

Equation:

$$\text{KE} = \frac{1}{2}mv^2,$$

is the energy associated with translational motion. Kinetic energy is a form of energy associated with the motion of a particle, single body, or system of objects moving together.

We are aware that it takes energy to get an object, like a car or the package in [\[link\]](#), up to speed, but it may be a bit surprising that kinetic energy is proportional to speed squared. This proportionality means, for example, that a car traveling at 100 km/h has four times the kinetic energy it has at 50

km/h, helping to explain why high-speed collisions are so devastating. We will now consider a series of examples to illustrate various aspects of work and energy.

Example:

Calculating the Kinetic Energy of a Package

Suppose a 30.0-kg package on the roller belt conveyor system in [\[link\]](#) is moving at 0.500 m/s. What is its kinetic energy?

Strategy

Because the mass m and speed v are given, the kinetic energy can be calculated from its definition as given in the equation $\text{KE} = \frac{1}{2}mv^2$.

Solution

The kinetic energy is given by

Equation:

$$\text{KE} = \frac{1}{2}mv^2.$$

Entering known values gives

Equation:

$$\text{KE} = 0.5(30.0 \text{ kg})(0.500 \text{ m/s})^2,$$

which yields

Equation:

$$\text{KE} = 3.75 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 3.75 \text{ J}.$$

Discussion

Note that the unit of kinetic energy is the joule, the same as the unit of work, as mentioned when work was first defined. It is also interesting that, although this is a fairly massive package, its kinetic energy is not large at this relatively low speed. This fact is consistent with the observation that people can move packages like this without exhausting themselves.

Example:**Determining the Work to Accelerate a Package**

Suppose that you push on the 30.0-kg package in [\[link\]](#) with a constant force of 120 N through a distance of 0.800 m, and that the opposing friction force averages 5.00 N.

(a) Calculate the net work done on the package. (b) Solve the same problem as in part (a), this time by finding the work done by each force that contributes to the net force.

Strategy and Concept for (a)

This is a motion in one dimension problem, because the downward force (from the weight of the package) and the normal force have equal magnitude and opposite direction, so that they cancel in calculating the net force, while the applied force, friction, and the displacement are all horizontal. (See [\[link\]](#).) As expected, the net work is the net force times distance.

Solution for (a)

The net force is the push force minus friction, or

$F_{\text{net}} = 120 \text{ N} - 5.00 \text{ N} = 115 \text{ N}$. Thus the net work is

Equation:

$$\begin{aligned} W_{\text{net}} &= F_{\text{net}}d = (115 \text{ N})(0.800 \text{ m}) \\ &= 92.0 \text{ N} \cdot \text{m} = 92.0 \text{ J.} \end{aligned}$$

Discussion for (a)

This value is the net work done on the package. The person actually does more work than this, because friction opposes the motion. Friction does negative work and removes some of the energy the person expends and converts it to thermal energy. The net work equals the sum of the work done by each individual force.

Strategy and Concept for (b)

The forces acting on the package are gravity, the normal force, the force of friction, and the applied force. The normal force and force of gravity are each perpendicular to the displacement, and therefore do no work.

Solution for (b)

The applied force does work.

Equation:

$$\begin{aligned}
 W_{\text{app}} &= F_{\text{app}} d \cos(0^\circ) = F_{\text{app}} d \\
 &= (120 \text{ N})(0.800 \text{ m}) \\
 &= 96.0 \text{ J}
 \end{aligned}$$

The friction force and displacement are in opposite directions, so that $\theta = 180^\circ$, and the work done by friction is

Equation:

$$\begin{aligned}
 W_{\text{fr}} &= F_{\text{fr}} d \cos(180^\circ) = -F_{\text{fr}} d \\
 &= -(5.00 \text{ N})(0.800 \text{ m}) \\
 &= -4.00 \text{ J}.
 \end{aligned}$$

So the amounts of work done by gravity, by the normal force, by the applied force, and by friction are, respectively,

Equation:

$$\begin{aligned}
 W_{\text{gr}} &= 0, \\
 W_{\text{N}} &= 0, \\
 W_{\text{app}} &= 96.0 \text{ J}, \\
 W_{\text{fr}} &= -4.00 \text{ J}.
 \end{aligned}$$

The total work done as the sum of the work done by each force is then seen to be

Equation:

$$W_{\text{total}} = W_{\text{gr}} + W_{\text{N}} + W_{\text{app}} + W_{\text{fr}} = 92.0 \text{ J}.$$

Discussion for (b)

The calculated total work W_{total} as the sum of the work by each force agrees, as expected, with the work W_{net} done by the net force. The work done by a collection of forces acting on an object can be calculated by either approach.

Example:

Determining Speed from Work and Energy

Find the speed of the package in [\[link\]](#) at the end of the push, using work and energy concepts.

Strategy

Here the work-energy theorem can be used, because we have just calculated the net work, W_{net} , and the initial kinetic energy, $\frac{1}{2}mv_0^2$. These calculations allow us to find the final kinetic energy, $\frac{1}{2}mv^2$, and thus the final speed v .

Solution

The work-energy theorem in equation form is

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

Solving for $\frac{1}{2}mv^2$ gives

Equation:

$$\frac{1}{2}mv^2 = W_{\text{net}} + \frac{1}{2}mv_0^2.$$

Thus,

Equation:

$$\frac{1}{2}mv^2 = 92.0 \text{ J} + 3.75 \text{ J} = 95.75 \text{ J}.$$

Solving for the final speed as requested and entering known values gives

Equation:

$$\begin{aligned} v &= \sqrt{\frac{2(95.75 \text{ J})}{m}} = \sqrt{\frac{191.5 \text{ kg}\cdot\text{m}^2/\text{s}^2}{30.0 \text{ kg}}} \\ &= 2.53 \text{ m/s}. \end{aligned}$$

Discussion

Using work and energy, we not only arrive at an answer, we see that the final kinetic energy is the sum of the initial kinetic energy and the net work

done on the package. This means that the work indeed adds to the energy of the package.

Example:**Work and Energy Can Reveal Distance, Too**

How far does the package in [\[link\]](#) coast after the push, assuming friction remains constant? Use work and energy considerations.

Strategy

We know that once the person stops pushing, friction will bring the package to rest. In terms of energy, friction does negative work until it has removed all of the package's kinetic energy. The work done by friction is the force of friction times the distance traveled times the cosine of the angle between the friction force and displacement; hence, this gives us a way of finding the distance traveled after the person stops pushing.

Solution

The normal force and force of gravity cancel in calculating the net force. The horizontal friction force is then the net force, and it acts opposite to the displacement, so $\theta = 180^\circ$. To reduce the kinetic energy of the package to zero, the work W_{fr} by friction must be minus the kinetic energy that the package started with plus what the package accumulated due to the pushing. Thus $W_{\text{fr}} = -95.75 \text{ J}$. Furthermore, $W_{\text{fr}} = f d' \cos \theta = -f d'$, where d' is the distance it takes to stop. Thus,

Equation:

$$d' = -\frac{W_{\text{fr}}}{f} = -\frac{-95.75 \text{ J}}{5.00 \text{ N}},$$

and so

Equation:

$$d' = 19.2 \text{ m}.$$

Discussion

This is a reasonable distance for a package to coast on a relatively friction-free conveyor system. Note that the work done by friction is negative (the

force is in the opposite direction of motion), so it removes the kinetic energy.

Some of the examples in this section can be solved without considering energy, but at the expense of missing out on gaining insights about what work and energy are doing in this situation. On the whole, solutions involving energy are generally shorter and easier than those using kinematics and dynamics alone.

Section Summary

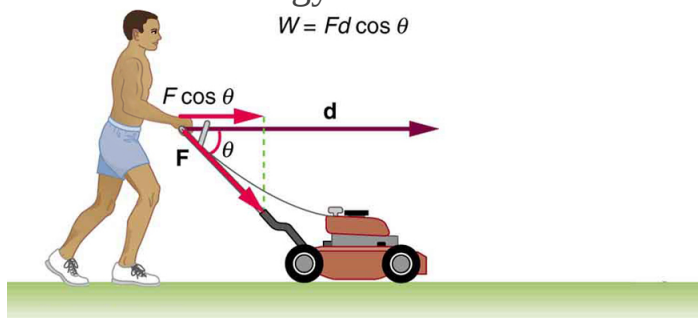
- The net work W_{net} is the work done by the net force acting on an object.
- Work done on an object transfers energy to the object.
- The translational kinetic energy of an object of mass m moving at speed v is $\text{KE} = \frac{1}{2}mv^2$.
- The work-energy theorem states that the net work W_{net} on a system changes its kinetic energy, $W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$.

Conceptual Questions

Exercise:

Problem:

The person in [\[link\]](#) does work on the lawn mower. Under what conditions would the mower gain energy? Under what conditions would it lose energy?



Exercise:**Problem:**

Work done on a system puts energy into it. Work done by a system removes energy from it. Give an example for each statement.

Exercise:**Problem:**

When solving for speed in [\[link\]](#), we kept only the positive root. Why?

Problems & Exercises**Exercise:****Problem:**

Compare the kinetic energy of a 20,000-kg truck moving at 110 km/h with that of an 80.0-kg astronaut in orbit moving at 27,500 km/h.

Solution:

1/250

Exercise:**Problem:**

(a) How fast must a 3000-kg elephant move to have the same kinetic energy as a 65.0-kg sprinter running at 10.0 m/s? (b) Discuss how the larger energies needed for the movement of larger animals would relate to metabolic rates.

Exercise:**Problem:**

Confirm the value given for the kinetic energy of an aircraft carrier in [\[link\]](#). You will need to look up the definition of a nautical mile (1 knot = 1 nautical mile/h).

Solution:

$$1.1 \times 10^{10} \text{ J}$$

Exercise:**Problem:**

(a) Calculate the force needed to bring a 950-kg car to rest from a speed of 90.0 km/h in a distance of 120 m (a fairly typical distance for a non-panic stop). (b) Suppose instead the car hits a concrete abutment at full speed and is brought to a stop in 2.00 m. Calculate the force exerted on the car and compare it with the force found in part (a).

Exercise:**Problem:**

A car's bumper is designed to withstand a 4.0-km/h (1.1-m/s) collision with an immovable object without damage to the body of the car. The bumper cushions the shock by absorbing the force over a distance. Calculate the magnitude of the average force on a bumper that collapses 0.200 m while bringing a 900-kg car to rest from an initial speed of 1.1 m/s.

Solution:

$$2.8 \times 10^3 \text{ N}$$

Exercise:

Problem:

Boxing gloves are padded to lessen the force of a blow. (a) Calculate the force exerted by a boxing glove on an opponent's face, if the glove and face compress 7.50 cm during a blow in which the 7.00-kg arm and glove are brought to rest from an initial speed of 10.0 m/s. (b) Calculate the force exerted by an identical blow in the gory old days when no gloves were used and the knuckles and face would compress only 2.00 cm. (c) Discuss the magnitude of the force with glove on. Does it seem high enough to cause damage even though it is lower than the force with no glove?

Exercise:**Problem:**

Using energy considerations, calculate the average force a 60.0-kg sprinter exerts backward on the track to accelerate from 2.00 to 8.00 m/s in a distance of 25.0 m, if he encounters a headwind that exerts an average force of 30.0 N against him.

Solution:

102 N

Glossary

net work

work done by the net force, or vector sum of all the forces, acting on an object

work-energy theorem

the result, based on Newton's laws, that the net work done on an object is equal to its change in kinetic energy

kinetic energy

the energy an object has by reason of its motion, equal to $\frac{1}{2}mv^2$ for the translational (i.e., non-rotational) motion of an object of mass m moving at speed v

Gravitational Potential Energy

- Explain gravitational potential energy in terms of work done against gravity.
- Show that the gravitational potential energy of an object of mass m at height h on Earth is given by $PE_g = mgh$.
- Show how knowledge of the potential energy as a function of position can be used to simplify calculations and explain physical phenomena.

Work Done Against Gravity

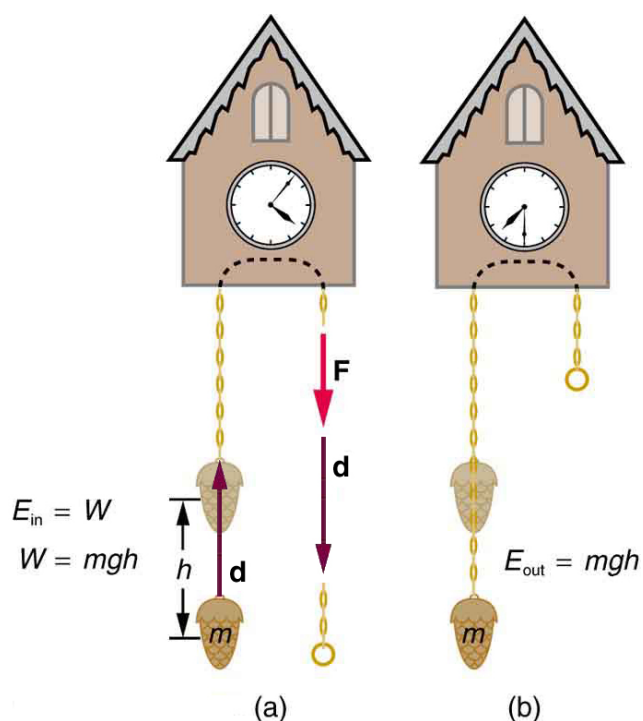
Climbing stairs and lifting objects is work in both the scientific and everyday sense—it is work done against the gravitational force. When there is work, there is a transformation of energy. The work done against the gravitational force goes into an important form of stored energy that we will explore in this section.

Let us calculate the work done in lifting an object of mass m through a height h , such as in [\[link\]](#). If the object is lifted straight up at constant speed, then the force needed to lift it is equal to its weight mg . The work done on the mass is then $W = Fd = mgh$. We define this to be the **gravitational potential energy** (PE_g) put into (or gained by) the object-Earth system. This energy is associated with the state of separation between two objects that attract each other by the gravitational force. For convenience, we refer to this as the PE_g gained by the object, recognizing that this is energy stored in the gravitational field of Earth. Why do we use the word “system”? Potential energy is a property of a system rather than of a single object—due to its physical position. An object’s gravitational potential is due to its position relative to the surroundings within the Earth-object system. The force applied to the object is an external force, from outside the system. When it does positive work it increases the gravitational potential energy of the system. Because gravitational potential energy depends on relative position, we need a reference level at which to set the potential energy equal to 0. We usually choose this point to be Earth’s surface, but this point is arbitrary; what is important is the *difference* in gravitational potential energy, because this difference is what relates to the work done. The difference in gravitational potential energy of an object (in the Earth-object system) between two rungs of a ladder will be the same for the first two rungs as for the last two rungs.

Converting Between Potential Energy and Kinetic Energy

Gravitational potential energy may be converted to other forms of energy, such as kinetic energy. If we release the mass, gravitational force will do an amount of work

equal to mgh on it, thereby increasing its kinetic energy by that same amount (by the work-energy theorem). We will find it more useful to consider just the conversion of PE_g to KE without explicitly considering the intermediate step of work. (See [\[link\]](#).) This shortcut makes it is easier to solve problems using energy (if possible) rather than explicitly using forces.



(a) The work done to lift the weight is stored in the mass-Earth system as gravitational potential energy. (b) As the weight moves downward, this gravitational potential energy is transferred to the cuckoo clock.

More precisely, we define the *change* in gravitational potential energy ΔPE_g to be
Equation:

$$\Delta PE_g = mgh,$$

where, for simplicity, we denote the change in height by h rather than the usual Δh . Note that h is positive when the final height is greater than the initial height, and vice versa. For example, if a 0.500-kg mass hung from a cuckoo clock is raised 1.00 m, then its change in gravitational potential energy is

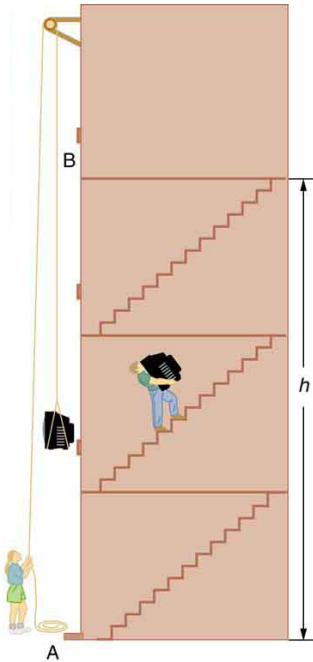
Equation:

$$\begin{aligned} mgh &= (0.500 \text{ kg}) (9.80 \text{ m/s}^2) (1.00 \text{ m}) \\ &= 4.90 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 4.90 \text{ J}. \end{aligned}$$

Note that the units of gravitational potential energy turn out to be joules, the same as for work and other forms of energy. As the clock runs, the mass is lowered. We can think of the mass as gradually giving up its 4.90 J of gravitational potential energy, *without directly considering the force of gravity that does the work*.

Using Potential Energy to Simplify Calculations

The equation $\Delta \text{PE}_g = mgh$ applies for any path that has a change in height of h , not just when the mass is lifted straight up. (See [\[link\]](#).) It is much easier to calculate mgh (a simple multiplication) than it is to calculate the work done along a complicated path. The idea of gravitational potential energy has the double advantage that it is very broadly applicable and it makes calculations easier. From now on, we will consider that any change in vertical position h of a mass m is accompanied by a change in gravitational potential energy mgh , and we will avoid the equivalent but more difficult task of calculating work done by or against the gravitational force.



The change in
gravitational
potential energy
(ΔPE_g)
between points
A and B is
independent of
the path.

$\Delta PE_g = mgh$
for any path
between the two
points. Gravity
is one of a small
class of forces
where the work
done by or
against the force
depends only on
the starting and
ending points,
not on the path
between them.

Example:**The Force to Stop Falling**

A 60.0-kg person jumps onto the floor from a height of 3.00 m. If he lands stiffly (with his knee joints compressing by 0.500 cm), calculate the force on the knee joints.

Strategy

This person's energy is brought to zero in this situation by the work done on him by the floor as he stops. The initial PE_g is transformed into KE as he falls. The work done by the floor reduces this kinetic energy to zero.

Solution

The work done on the person by the floor as he stops is given by

Equation:

$$W = Fd \cos \theta = -Fd,$$

with a minus sign because the displacement while stopping and the force from floor are in opposite directions ($\cos \theta = \cos 180^\circ = -1$). The floor removes energy from the system, so it does negative work.

The kinetic energy the person has upon reaching the floor is the amount of potential energy lost by falling through height h :

Equation:

$$KE = -\Delta PE_g = -mgh,$$

The distance d that the person's knees bend is much smaller than the height h of the fall, so the additional change in gravitational potential energy during the knee bend is ignored.

The work W done by the floor on the person stops the person and brings the person's kinetic energy to zero:

Equation:

$$W = -KE = mgh.$$

Combining this equation with the expression for W gives

Equation:

$$-Fd = mgh.$$

Recalling that h is negative because the person fell *down*, the force on the knee joints is given by

Equation:

$$F = -\frac{mgh}{d} = -\frac{(60.0 \text{ kg})(9.80 \text{ m/s}^2)(-3.00 \text{ m})}{5.00 \times 10^{-3} \text{ m}} = 3.53 \times 10^5 \text{ N}.$$

Discussion

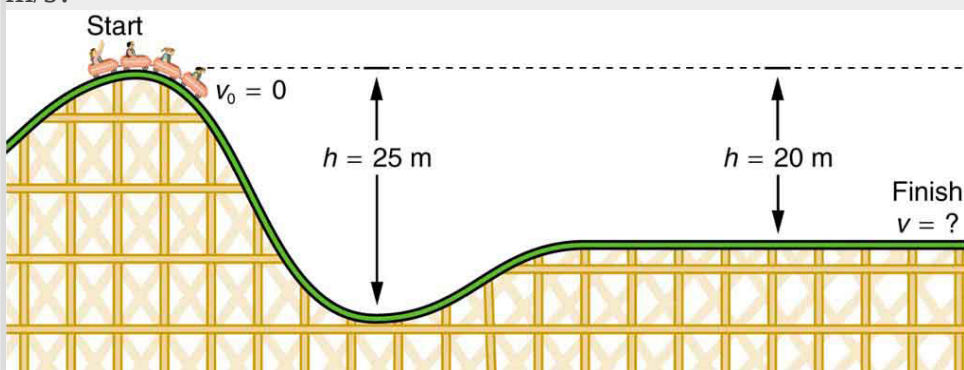
Such a large force (500 times more than the person's weight) over the short impact time is enough to break bones. A much better way to cushion the shock is by bending the legs or rolling on the ground, increasing the time over which the force acts. A bending motion of 0.5 m this way yields a force 100 times smaller than in the example. A kangaroo's hopping shows this method in action. The kangaroo is the only large animal to use hopping for locomotion, but the shock in hopping is cushioned by the bending of its hind legs in each jump. (See [\[link\]](#).)



The work done by the ground upon the kangaroo reduces its kinetic energy to zero as it lands. However, by applying the force of the ground on the hind legs over a longer distance, the impact on the bones is reduced.
(credit: Chris Samuel, Flickr)

Example:**Finding the Speed of a Roller Coaster from its Height**

(a) What is the final speed of the roller coaster shown in [\[link\]](#) if it starts from rest at the top of the 20.0 m hill and work done by frictional forces is negligible? (b) What is its final speed (again assuming negligible friction) if its initial speed is 5.00 m/s?



The speed of a roller coaster increases as gravity pulls it downhill and is greatest at its lowest point. Viewed in terms of energy, the roller-coaster-Earth system's gravitational potential energy is converted to kinetic energy. If work done by friction is negligible, all ΔPE_g is converted to KE.

Strategy

The roller coaster loses potential energy as it goes downhill. We neglect friction, so that the remaining force exerted by the track is the normal force, which is perpendicular to the direction of motion and does no work. The net work on the roller coaster is then done by gravity alone. The *loss* of gravitational potential energy from moving *downward* through a distance h equals the *gain* in kinetic energy. This can be written in equation form as $-\Delta PE_g = \Delta KE$. Using the equations for PE_g and KE, we can solve for the final speed v , which is the desired quantity.

Solution for (a)

Here the initial kinetic energy is zero, so that $\Delta KE = \frac{1}{2}mv^2$. The equation for change in potential energy states that $\Delta PE_g = mgh$. Since h is negative in this case, we will rewrite this as $\Delta PE_g = -mg |h|$ to show the minus sign clearly. Thus,

Equation:

$$-\Delta\text{PE}_g = \Delta\text{KE}$$

becomes

Equation:

$$mg | h | = \frac{1}{2}mv^2.$$

Solving for v , we find that mass cancels and that

Equation:

$$v = \sqrt{2g | h |}.$$

Substituting known values,

Equation:

$$\begin{aligned} v &= \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m})} \\ &= 19.8 \text{ m/s.} \end{aligned}$$

Solution for (b)

Again $-\Delta\text{PE}_g = \Delta\text{KE}$. In this case there is initial kinetic energy, so

$\Delta\text{KE} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$. Thus,

Equation:

$$mg | h | = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

Rearranging gives

Equation:

$$\frac{1}{2}mv^2 = mg | h | + \frac{1}{2}mv_0^2.$$

This means that the final kinetic energy is the sum of the initial kinetic energy and the gravitational potential energy. Mass again cancels, and

Equation:

$$v = \sqrt{2g | h | + v_0^2}.$$

This equation is very similar to the kinematics equation $v = \sqrt{v_0^2 + 2ad}$, but it is more general—the kinematics equation is valid only for constant acceleration, whereas our equation above is valid for any path regardless of whether the object moves with a constant acceleration. Now, substituting known values gives

Equation:

$$\begin{aligned} v &= \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m}) + (5.00 \text{ m/s})^2} \\ &= 20.4 \text{ m/s.} \end{aligned}$$

Discussion and Implications

First, note that mass cancels. This is quite consistent with observations made in [Falling Objects](#) that all objects fall at the same rate if friction is negligible. Second, only the speed of the roller coaster is considered; there is no information about its direction at any point. This reveals another general truth. When friction is negligible, the speed of a falling body depends only on its initial speed and height, and not on its mass or the path taken. For example, the roller coaster will have the same final speed whether it falls 20.0 m straight down or takes a more complicated path like the one in the figure. Third, and perhaps unexpectedly, the final speed in part (b) is greater than in part (a), but by far less than 5.00 m/s. Finally, note that speed can be found at *any* height along the way by simply using the appropriate value of h at the point of interest.

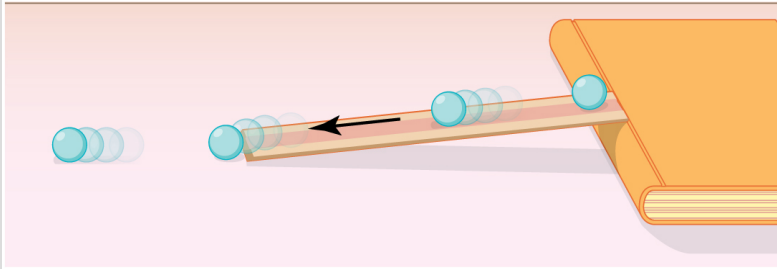
We have seen that work done by or against the gravitational force depends only on the starting and ending points, and not on the path between, allowing us to define the simplifying concept of gravitational potential energy. We can do the same thing for a few other forces, and we will see that this leads to a formal definition of the law of conservation of energy.

Note:

Making Connections: Take-Home Investigation—Converting Potential to Kinetic Energy

One can study the conversion of gravitational potential energy into kinetic energy in this experiment. On a smooth, level surface, use a ruler of the kind that has a groove running along its length and a book to make an incline (see [link](#)). Place a marble at the 10-cm position on the ruler and let it roll down the ruler. When it hits the level surface, measure the time it takes to roll one meter. Now place the marble

at the 20-cm and the 30-cm positions and again measure the times it takes to roll 1 m on the level surface. Find the velocity of the marble on the level surface for all three positions. Plot velocity squared versus the distance traveled by the marble. What is the shape of each plot? If the shape is a straight line, the plot shows that the marble's kinetic energy at the bottom is proportional to its potential energy at the release point.



A marble rolls down a ruler, and its speed on the level surface is measured.

Section Summary

- Work done against gravity in lifting an object becomes potential energy of the object-Earth system.
- The change in gravitational potential energy, ΔPE_g , is $\Delta PE_g = mgh$, with h being the increase in height and g the acceleration due to gravity.
- The gravitational potential energy of an object near Earth's surface is due to its position in the mass-Earth system. Only differences in gravitational potential energy, ΔPE_g , have physical significance.
- As an object descends without friction, its gravitational potential energy changes into kinetic energy corresponding to increasing speed, so that $\Delta KE = -\Delta PE_g$.

Conceptual Questions

Exercise:

Problem:

In [\[link\]](#), we calculated the final speed of a roller coaster that descended 20 m in height and had an initial speed of 5 m/s downhill. Suppose the roller coaster had had an initial speed of 5 m/s *uphill* instead, and it coasted uphill, stopped, and then rolled back down to a final point 20 m below the start. We would find in that case that its final speed is the same as its initial speed. Explain in terms of conservation of energy.

Exercise:**Problem:**

Does the work you do on a book when you lift it onto a shelf depend on the path taken? On the time taken? On the height of the shelf? On the mass of the book?

Problems & Exercises**Exercise:****Problem:**

A hydroelectric power facility (see [\[link\]](#)) converts the gravitational potential energy of water behind a dam to electric energy. (a) What is the gravitational potential energy relative to the generators of a lake of volume 50.0 km^3 (mass = $5.00 \times 10^{13} \text{ kg}$), given that the lake has an average height of 40.0 m above the generators? (b) Compare this with the energy stored in a 9-megaton fusion bomb.



Hydroelectric facility (credit: Denis

Solution:

(a) $1.96 \times 10^{16} \text{ J}$

(b) The ratio of gravitational potential energy in the lake to the energy stored in the bomb is 0.52. That is, the energy stored in the lake is approximately half that in a 9-megaton fusion bomb.

Exercise:

Problem:

(a) How much gravitational potential energy (relative to the ground on which it is built) is stored in the Great Pyramid of Cheops, given that its mass is about $7 \times 10^9 \text{ kg}$ and its center of mass is 36.5 m above the surrounding ground? (b) How does this energy compare with the daily food intake of a person?

Exercise:

Problem:

Suppose a 350-g kookaburra (a large kingfisher bird) picks up a 75-g snake and raises it 2.5 m from the ground to a branch. (a) How much work did the bird do on the snake? (b) How much work did it do to raise its own center of mass to the branch?

Solution:

(a) 1.8 J

(b) 8.6 J

Exercise:

Problem:

In [\[link\]](#), we found that the speed of a roller coaster that had descended 20.0 m was only slightly greater when it had an initial speed of 5.00 m/s than when it started from rest. This implies that $\Delta PE \gg KE_i$. Confirm this statement by taking the ratio of ΔPE to KE_i . (Note that mass cancels.)

Exercise:

Problem:

A 100-g toy car is propelled by a compressed spring that starts it moving. The car follows the curved track in [\[link\]](#). Show that the final speed of the toy car is 0.687 m/s if its initial speed is 2.00 m/s and it coasts up the frictionless slope, gaining 0.180 m in altitude.



A toy car moves up a sloped track.
(credit: Leszek Leszczynski, Flickr)

Solution:**Equation:**

$$v_f = \sqrt{2gh + v_0^2} = \sqrt{2(9.80 \text{ m/s}^2)(-0.180 \text{ m}) + (2.00 \text{ m/s})^2} = 0.687 \text{ m/s}$$

Exercise:**Problem:**

In a downhill ski race, surprisingly, little advantage is gained by getting a running start. (This is because the initial kinetic energy is small compared with the gain in gravitational potential energy on even small hills.) To demonstrate this, find the final speed and the time taken for a skier who skies 70.0 m along a 30° slope neglecting friction: (a) Starting from rest. (b) Starting with an initial speed of 2.50 m/s. (c) Does the answer surprise you? Discuss why it is still advantageous to get a running start in very competitive events.

Glossary

gravitational potential energy

the energy an object has due to its position in a gravitational field

Conservative Forces and Potential Energy

- Define conservative force, potential energy, and mechanical energy.
- Explain the potential energy of a spring in terms of its compression when Hooke's law applies.
- Use the work-energy theorem to show how having only conservative forces implies conservation of mechanical energy.

Potential Energy and Conservative Forces

Work is done by a force, and some forces, such as weight, have special characteristics. A **conservative force** is one, like the gravitational force, for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken. We can define a **potential energy** (PE) for any conservative force, just as we did for the gravitational force. For example, when you wind up a toy, an egg timer, or an old-fashioned watch, you do work against its spring and store energy in it. (We treat these springs as ideal, in that we assume there is no friction and no production of thermal energy.) This stored energy is recoverable as work, and it is useful to think of it as potential energy contained in the spring. Indeed, the reason that the spring has this characteristic is that its force is *conservative*. That is, a conservative force results in stored or potential energy. Gravitational potential energy is one example, as is the energy stored in a spring. We will also see how conservative forces are related to the conservation of energy.

Note:

Potential Energy and Conservative Forces

Potential energy is the energy a system has due to position, shape, or configuration. It is stored energy that is completely recoverable.

A conservative force is one for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken.

We can define a potential energy (PE) for any conservative force. The work done against a conservative force to reach a final configuration

depends on the configuration, not the path followed, and is the potential energy added.

Potential Energy of a Spring

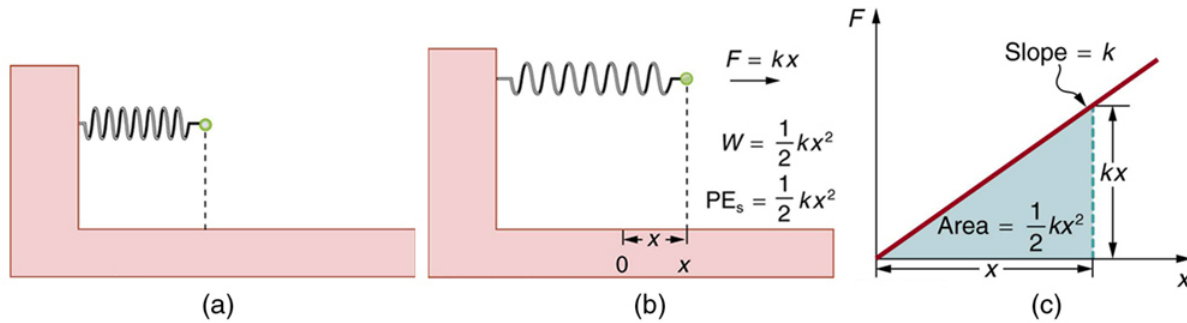
First, let us obtain an expression for the potential energy stored in a spring (PE_s). We calculate the work done to stretch or compress a spring that obeys Hooke's law. (Hooke's law was examined in [Elasticity: Stress and Strain](#), and states that the magnitude of force F on the spring and the resulting deformation ΔL are proportional, $F = k\Delta L$.) (See [\[link\]](#).) For our spring, we will replace ΔL (the amount of deformation produced by a force F) by the distance x that the spring is stretched or compressed along its length. So the force needed to stretch the spring has magnitude $F = kx$, where k is the spring's force constant. The force increases linearly from 0 at the start to kx in the fully stretched position. The average force is $kx/2$. Thus the work done in stretching or compressing the spring is

$W_s = Fd = \left(\frac{kx}{2}\right)x = \frac{1}{2}kx^2$. Alternatively, we noted in [Kinetic Energy and the Work-Energy Theorem](#) that the area under a graph of F vs. x is the work done by the force. In [\[link\]](#)(c) we see that this area is also $\frac{1}{2}kx^2$. We therefore define the **potential energy of a spring**, PE_s , to be

Equation:

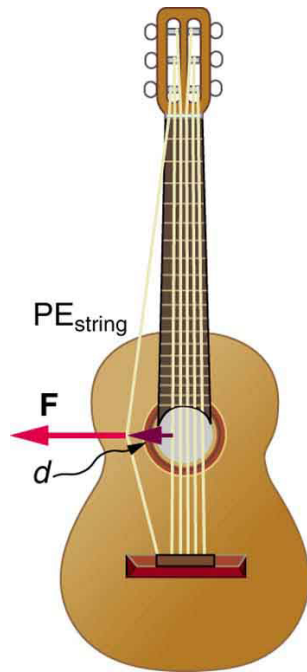
$$PE_s = \frac{1}{2}kx^2,$$

where k is the spring's force constant and x is the displacement from its undeformed position. The potential energy represents the work done *on* the spring and the energy stored in it as a result of stretching or compressing it a distance x . The potential energy of the spring PE_s does not depend on the path taken; it depends only on the stretch or squeeze x in the final configuration.



- (a) An undeformed spring has no PE_s stored in it. (b) The force needed to stretch (or compress) the spring a distance x has a magnitude $F = kx$, and the work done to stretch (or compress) it is $\frac{1}{2} kx^2$. Because the force is conservative, this work is stored as potential energy (PE_s) in the spring, and it can be fully recovered. (c) A graph of F vs. x has a slope of k , and the area under the graph is $\frac{1}{2} kx^2$. Thus the work done or potential energy stored is $\frac{1}{2} kx^2$.

The equation $PE_s = \frac{1}{2} kx^2$ has general validity beyond the special case for which it was derived. Potential energy can be stored in any elastic medium by deforming it. Indeed, the general definition of **potential energy** is energy due to position, shape, or configuration. For shape or position deformations, stored energy is $PE_s = \frac{1}{2} kx^2$, where k is the force constant of the particular system and x is its deformation. Another example is seen in [\[link\]](#) for a guitar string.



Work is done
to deform the
guitar string,
giving it
potential
energy.

When
released, the
potential
energy is
converted to
kinetic
energy and
back to
potential as
the string
oscillates
back and
forth. A very
small
fraction is
dissipated as

sound
energy,
slowly
removing
energy from
the string.

Conservation of Mechanical Energy

Let us now consider what form the work-energy theorem takes when only conservative forces are involved. This will lead us to the conservation of energy principle. The work-energy theorem states that the net work done by all forces acting on a system equals its change in kinetic energy. In equation form, this is

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2 = \Delta\text{KE}.$$

If only conservative forces act, then

Equation:

$$W_{\text{net}} = W_{\text{c}},$$

where W_{c} is the total work done by all conservative forces. Thus,

Equation:

$$W_{\text{c}} = \Delta\text{KE}.$$

Now, if the conservative force, such as the gravitational force or a spring force, does work, the system loses potential energy. That is, $W_{\text{c}} = -\Delta\text{PE}$. Therefore,

Equation:

$$-\Delta PE = \Delta KE$$

or

Equation:

$$\Delta KE + \Delta PE = 0.$$

This equation means that the total kinetic and potential energy is constant for any process involving only conservative forces. That is,

Equation:

$$KE + PE = \text{constant}$$

or

(conservative forces only),

$$KE_i + PE_i = KE_f + PE_f$$

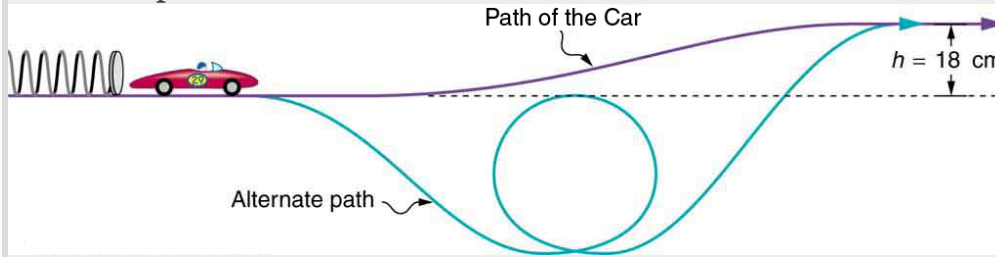
where i and f denote initial and final values. This equation is a form of the work-energy theorem for conservative forces; it is known as the **conservation of mechanical energy** principle. Remember that this applies to the extent that all the forces are conservative, so that friction is negligible. The total kinetic plus potential energy of a system is defined to be its **mechanical energy**, $(KE + PE)$. In a system that experiences only conservative forces, there is a potential energy associated with each force, and the energy only changes form between KE and the various types of PE, with the total energy remaining constant.

Example:

Using Conservation of Mechanical Energy to Calculate the Speed of a Toy Car

A 0.100-kg toy car is propelled by a compressed spring, as shown in [\[link\]](#). The car follows a track that rises 0.180 m above the starting point. The spring is compressed 4.00 cm and has a force constant of 250.0 N/m. Assuming work done by friction to be negligible, find (a) how fast the car

is going before it starts up the slope and (b) how fast it is going at the top of the slope.



A toy car is pushed by a compressed spring and coasts up a slope. Assuming negligible friction, the potential energy in the spring is first completely converted to kinetic energy, and then to a combination of kinetic and gravitational potential energy as the car rises. The details of the path are unimportant because all forces are conservative—the car would have the same final speed if it took the alternate path shown.

Strategy

The spring force and the gravitational force are conservative forces, so conservation of mechanical energy can be used. Thus,

Equation:

$$KE_i + PE_i = KE_f + PE_f$$

or

Equation:

$$\frac{1}{2}mv_i^2 + mgh_i + \frac{1}{2}kx_i^2 = \frac{1}{2}mv_f^2 + mgh_f + \frac{1}{2}kx_f^2,$$

where h is the height (vertical position) and x is the compression of the spring. This general statement looks complex but becomes much simpler when we start considering specific situations. First, we must identify the initial and final conditions in a problem; then, we enter them into the last equation to solve for an unknown.

Solution for (a)

This part of the problem is limited to conditions just before the car is released and just after it leaves the spring. Take the initial height to be zero, so that both h_i and h_f are zero. Furthermore, the initial speed v_i is zero and the final compression of the spring x_f is zero, and so several terms in the conservation of mechanical energy equation are zero and it simplifies to

Equation:

$$\frac{1}{2} k x_i^2 = \frac{1}{2} m v_f^2.$$

In other words, the initial potential energy in the spring is converted completely to kinetic energy in the absence of friction. Solving for the final speed and entering known values yields

Equation:

$$\begin{aligned} v_f &= \sqrt{\frac{k}{m}} x_i \\ &= \sqrt{\frac{250.0 \text{ N/m}}{0.100 \text{ kg}}} (0.0400 \text{ m}) \\ &= 2.00 \text{ m/s.} \end{aligned}$$

Solution for (b)

One method of finding the speed at the top of the slope is to consider conditions just before the car is released and just after it reaches the top of the slope, completely ignoring everything in between. Doing the same type of analysis to find which terms are zero, the conservation of mechanical energy becomes

Equation:

$$\frac{1}{2} k x_i^2 = \frac{1}{2} m v_f^2 + m g h_f.$$

This form of the equation means that the spring's initial potential energy is converted partly to gravitational potential energy and partly to kinetic energy. The final speed at the top of the slope will be less than at the bottom. Solving for v_f and substituting known values gives

Equation:

$$\begin{aligned}
 v_f &= \sqrt{\frac{kx_i^2}{m} - 2gh_f} \\
 &= \sqrt{\left(\frac{250.0 \text{ N/m}}{0.100 \text{ kg}}\right)(0.0400 \text{ m})^2 - 2(9.80 \text{ m/s}^2)(0.180 \text{ m})} \\
 &= 0.687 \text{ m/s.}
 \end{aligned}$$

Discussion

Another way to solve this problem is to realize that the car's kinetic energy before it goes up the slope is converted partly to potential energy—that is, to take the final conditions in part (a) to be the initial conditions in part (b).

Note that, for conservative forces, we do not directly calculate the work they do; rather, we consider their effects through their corresponding potential energies, just as we did in [\[link\]](#). Note also that we do not consider details of the path taken—only the starting and ending points are important (as long as the path is not impossible). This assumption is usually a tremendous simplification, because the path may be complicated and forces may vary along the way.

Note:

PhET Explorations: Energy Skate Park

Learn about conservation of energy with a skater dude! Build tracks, ramps and jumps for the skater and view the kinetic energy, potential energy and friction as he moves. You can also take the skater to different planets or even space!

https://phet.colorado.edu/sims/html/energy-skate-park-basics/latest/energy-skate-park-basics_en.html

Section Summary

- A conservative force is one for which work depends only on the starting and ending points of a motion, not on the path taken.
- We can define potential energy (PE) for any conservative force, just as we defined PE_g for the gravitational force.
- The potential energy of a spring is $PE_s = \frac{1}{2}kx^2$, where k is the spring's force constant and x is the displacement from its undeformed position.
- Mechanical energy is defined to be $KE + PE$ for a conservative force.
- When only conservative forces act on and within a system, the total mechanical energy is constant. In equation form,

Equation:

$$KE + PE = \text{constant}$$

or

$$KE_i + PE_i = KE_f + PE_f$$

where i and f denote initial and final values. This is known as the conservation of mechanical energy.

Conceptual Questions

Exercise:

Problem: What is a conservative force?

Exercise:

Problem:

The force exerted by a diving board is conservative, provided the internal friction is negligible. Assuming friction is negligible, describe changes in the potential energy of a diving board as a swimmer dives from it, starting just before the swimmer steps on the board until just after his feet leave it.

Exercise:

Problem:

Define mechanical energy. What is the relationship of mechanical energy to nonconservative forces? What happens to mechanical energy if only conservative forces act?

Exercise:**Problem:**

What is the relationship of potential energy to conservative force?

Problems & Exercises**Exercise:****Problem:**

A 5.00×10^5 -kg subway train is brought to a stop from a speed of 0.500 m/s in 0.400 m by a large spring bumper at the end of its track. What is the force constant k of the spring?

Solution:**Equation:**

$$7.81 \times 10^5 \text{ N/m}$$

Exercise:**Problem:**

A pogo stick has a spring with a force constant of 2.50×10^4 N/m, which can be compressed 12.0 cm. To what maximum height can a child jump on the stick using only the energy in the spring, if the child and stick have a total mass of 40.0 kg? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Energy](#).

Glossary

conservative force

a force that does the same work for any given initial and final configuration, regardless of the path followed

potential energy

energy due to position, shape, or configuration

potential energy of a spring

the stored energy of a spring as a function of its displacement; when Hooke's law applies, it is given by the expression $\frac{1}{2}kx^2$ where x is the distance the spring is compressed or extended and k is the spring constant

conservation of mechanical energy

the rule that the sum of the kinetic energies and potential energies remains constant if only conservative forces act on and within a system

mechanical energy

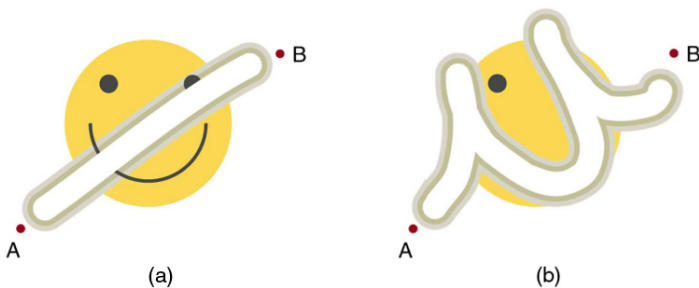
the sum of kinetic energy and potential energy

Nonconservative Forces

- Define nonconservative forces and explain how they affect mechanical energy.
- Show how the principle of conservation of energy can be applied by treating the conservative forces in terms of their potential energies and any nonconservative forces in terms of the work they do.

Nonconservative Forces and Friction

Forces are either conservative or nonconservative. Conservative forces were discussed in [Conservative Forces and Potential Energy](#). A **nonconservative force** is one for which work depends on the path taken. Friction is a good example of a nonconservative force. As illustrated in [\[link\]](#), work done against friction depends on the length of the path between the starting and ending points. Because of this dependence on path, there is no potential energy associated with nonconservative forces. An important characteristic is that the work done by a nonconservative force *adds or removes mechanical energy from a system*. **Friction**, for example, creates **thermal energy** that dissipates, removing energy from the system. Furthermore, even if the thermal energy is retained or captured, it cannot be fully converted back to work, so it is lost or not recoverable in that sense as well.

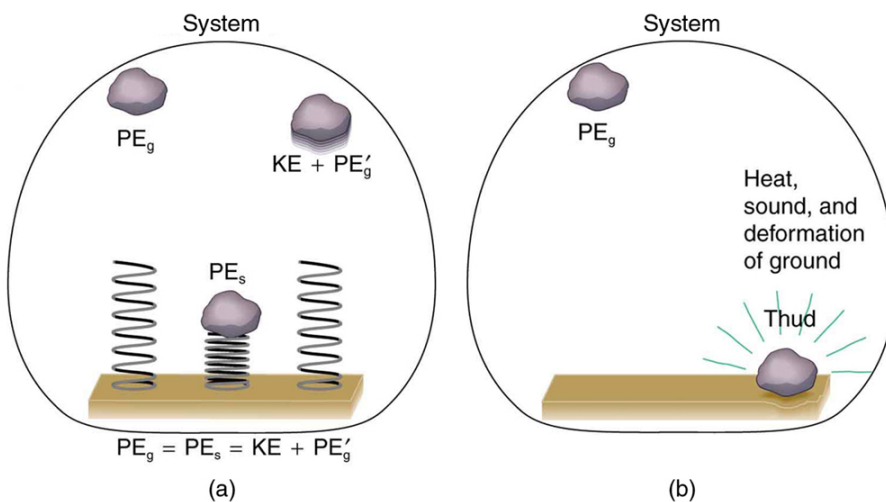


The amount of the happy face erased depends on the path taken by the eraser between points A and B, as does the work done against friction. Less work is done and less of the face

is erased for the path in (a) than for the path in (b). The force here is friction, and most of the work goes into thermal energy that subsequently leaves the system (the happy face plus the eraser). The energy expended cannot be fully recovered.

How Nonconservative Forces Affect Mechanical Energy

Mechanical energy may not be conserved when nonconservative forces act. For example, when a car is brought to a stop by friction on level ground, it loses kinetic energy, which is dissipated as thermal energy, reducing its mechanical energy. [\[link\]](#) compares the effects of conservative and nonconservative forces. We often choose to understand simpler systems such as that described in [\[link\]](#)(a) first before studying more complicated systems as in [\[link\]](#)(b).



Comparison of the effects of conservative and nonconservative forces on the mechanical energy of a system. (a) A system with only conservative

forces. When a rock is dropped onto a spring, its mechanical energy remains constant (neglecting air resistance) because the force in the spring is conservative. The spring can propel the rock back to its original height, where it once again has only potential energy due to gravity. (b) A system with nonconservative forces. When the same rock is dropped onto the ground, it is stopped by nonconservative forces that dissipate its mechanical energy as thermal energy, sound, and surface distortion. The rock has lost mechanical energy.

How the Work-Energy Theorem Applies

Now let us consider what form the work-energy theorem takes when both conservative and nonconservative forces act. We will see that the work done by nonconservative forces equals the change in the mechanical energy of a system. As noted in [Kinetic Energy and the Work-Energy Theorem](#), the work-energy theorem states that the net work on a system equals the change in its kinetic energy, or $W_{\text{net}} = \Delta\text{KE}$. The net work is the sum of the work by nonconservative forces plus the work by conservative forces. That is,

Equation:

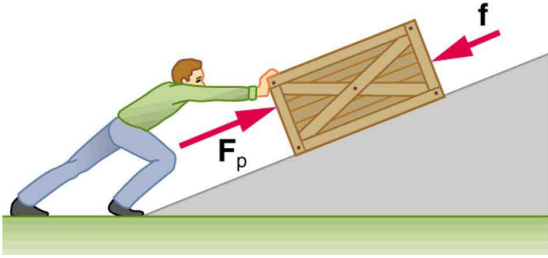
$$W_{\text{net}} = W_{\text{nc}} + W_{\text{c}},$$

so that

Equation:

$$W_{\text{nc}} + W_{\text{c}} = \Delta\text{KE},$$

where W_{nc} is the total work done by all nonconservative forces and W_{c} is the total work done by all conservative forces.



A person pushes a crate up a ramp, doing work on the crate. Friction and gravitational force (not shown) also do work on the crate; both forces oppose the person's push. As the crate is pushed up the ramp, it gains mechanical energy, implying that the work done by the person is greater than the work done by friction.

Consider [\[link\]](#), in which a person pushes a crate up a ramp and is opposed by friction. As in the previous section, we note that work done by a conservative force comes from a loss of gravitational potential energy, so that $W_c = -\Delta PE$. Substituting this equation into the previous one and solving for W_{nc} gives

Equation:

$$W_{nc} = \Delta KE + \Delta PE.$$

This equation means that the total mechanical energy ($KE + PE$) changes by exactly the amount of work done by nonconservative forces. In [\[link\]](#), this is the work done by the person minus the work done by friction. So even if energy is not conserved for the system of interest (such as the crate), we know that an equal amount of work was done to cause the change in total mechanical energy.

We rearrange $W_{\text{nc}} = \Delta\text{KE} + \Delta\text{PE}$ to obtain

Equation:

$$\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f.$$

This means that the amount of work done by nonconservative forces adds to the mechanical energy of a system. If W_{nc} is positive, then mechanical energy is increased, such as when the person pushes the crate up the ramp in [\[link\]](#). If W_{nc} is negative, then mechanical energy is decreased, such as when the rock hits the ground in [\[link\]](#)(b). If W_{nc} is zero, then mechanical energy is conserved, and nonconservative forces are balanced. For example, when you push a lawn mower at constant speed on level ground, your work done is removed by the work of friction, and the mower has a constant energy.

Applying Energy Conservation with Nonconservative Forces

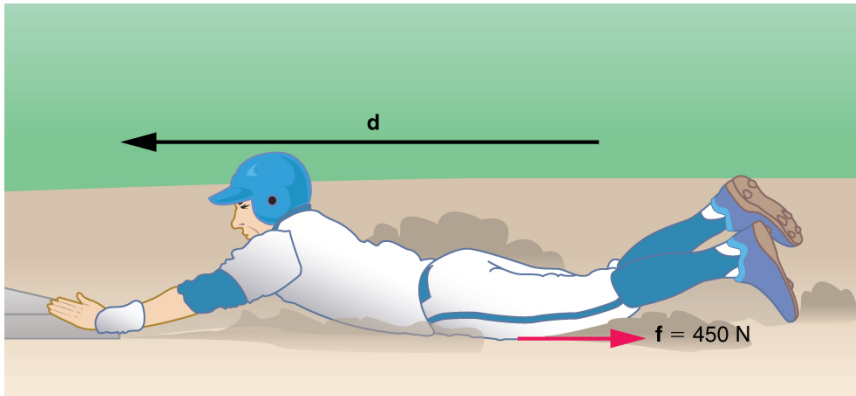
When no change in potential energy occurs, applying $\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f$ amounts to applying the work-energy theorem by setting the change in kinetic energy to be equal to the net work done on the system, which in the most general case includes both conservative and nonconservative forces. But when seeking instead to find a change in total mechanical energy in situations that involve changes in both potential and kinetic energy, the previous equation $\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f$ says that you can start by finding the change in mechanical energy that would have resulted from just the conservative forces, including the potential energy changes, and add to it the work done, with the proper sign, by any nonconservative forces involved.

Example:

Calculating Distance Traveled: How Far a Baseball Player Slides

Consider the situation shown in [\[link\]](#), where a baseball player slides to a stop on level ground. Using energy considerations, calculate the distance

the 65.0-kg baseball player slides, given that his initial speed is 6.00 m/s and the force of friction against him is a constant 450 N.



The baseball player slides to a stop in a distance d . In the process, friction removes the player's kinetic energy by doing an amount of work fd equal to the initial kinetic energy.

Strategy

Friction stops the player by converting his kinetic energy into other forms, including thermal energy. In terms of the work-energy theorem, the work done by friction, which is negative, is added to the initial kinetic energy to reduce it to zero. The work done by friction is negative, because \mathbf{f} is in the opposite direction of the motion (that is, $\theta = 180^\circ$, and so $\cos \theta = -1$). Thus $W_{\text{nc}} = -fd$. The equation simplifies to

Equation:

$$\frac{1}{2}mv_i^2 - fd = 0$$

or

Equation:

$$fd = \frac{1}{2}mv_i^2.$$

This equation can now be solved for the distance d .

Solution

Solving the previous equation for d and substituting known values yields
Equation:

$$\begin{aligned}d &= \frac{mv_i^2}{2f} \\&= \frac{(65.0 \text{ kg})(6.00 \text{ m/s})^2}{(2)(450 \text{ N})} \\&= 2.60 \text{ m.}\end{aligned}$$

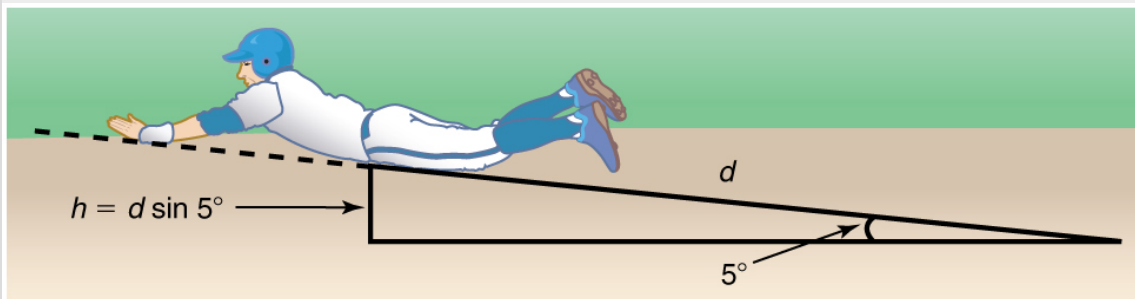
Discussion

The most important point of this example is that the amount of nonconservative work equals the change in mechanical energy. For example, you must work harder to stop a truck, with its large mechanical energy, than to stop a mosquito.

Example:

Calculating Distance Traveled: Sliding Up an Incline

Suppose that the player from [\[link\]](#) is running up a hill having a 5.00° incline upward with a surface similar to that in the baseball stadium. The player slides with the same initial speed, and the frictional force is still 450 N. Determine how far he slides.



The same baseball player slides to a stop on a 5.00° slope.

Strategy

In this case, the work done by the nonconservative friction force on the player reduces the mechanical energy he has from his kinetic energy at zero height, to the final mechanical energy he has by moving through

distance d to reach height h along the hill, with $h = d \sin 5.00^\circ$. This is expressed by the equation

Equation:

$$\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f.$$

Solution

The work done by friction is again $W_{\text{nc}} = -fd$; initially the potential energy is $\text{PE}_i = mg \cdot 0 = 0$ and the kinetic energy is $\text{KE}_i = \frac{1}{2}mv_i^2$; the final energy contributions are $\text{KE}_f = 0$ for the kinetic energy and $\text{PE}_f = mgh = mgd \sin \theta$ for the potential energy.

Substituting these values gives

Equation:

$$\frac{1}{2}mv_i^2 + 0 + (-fd) = 0 + mgd \sin \theta.$$

Solve this for d to obtain

Equation:

$$\begin{aligned} d &= \frac{(\frac{1}{2})mv_i^2}{f + mg \sin \theta} \\ &= \frac{(0.5)(65.0 \text{ kg})(6.00 \text{ m/s})^2}{450 \text{ N} + (65.0 \text{ kg})(9.80 \text{ m/s}^2) \sin (5.00^\circ)} \\ &= 2.31 \text{ m.} \end{aligned}$$

Discussion

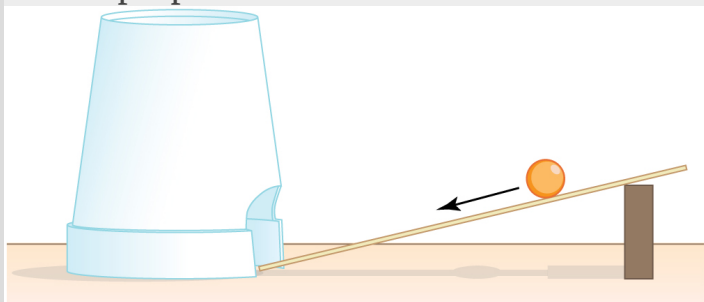
As might have been expected, the player slides a shorter distance by sliding uphill. Note that the problem could also have been solved in terms of the forces directly and the work energy theorem, instead of using the potential energy. This method would have required combining the normal force and force of gravity vectors, which no longer cancel each other because they point in different directions, and friction, to find the net force. You could then use the net force and the net work to find the distance d that reduces the kinetic energy to zero. By applying conservation of energy and using the potential energy instead, we need only consider the gravitational potential energy mgh , without combining and resolving force vectors. This simplifies the solution considerably.

Note:**Making Connections: Take-Home Investigation—Determining Friction from the Stopping Distance**

This experiment involves the conversion of gravitational potential energy into thermal energy. Use the ruler, book, and marble from [Take-Home Investigation—Converting Potential to Kinetic Energy](#). In addition, you will need a foam cup with a small hole in the side, as shown in [\[link\]](#). From the 10-cm position on the ruler, let the marble roll into the cup positioned at the bottom of the ruler. Measure the distance d the cup moves before stopping. What forces caused it to stop? What happened to the kinetic energy of the marble at the bottom of the ruler? Next, place the marble at the 20-cm and the 30-cm positions and again measure the distance the cup moves after the marble enters it. Plot the distance the cup moves versus the initial marble position on the ruler. Is this relationship linear?

With some simple assumptions, you can use these data to find the coefficient of kinetic friction μ_k of the cup on the table. The force of friction f on the cup is $\mu_k N$, where the normal force N is just the weight of the cup plus the marble. The normal force and force of gravity do no work because they are perpendicular to the displacement of the cup, which moves horizontally. The work done by friction is fd . You will need the mass of the marble as well to calculate its initial kinetic energy.

It is interesting to do the above experiment also with a steel marble (or ball bearing). Releasing it from the same positions on the ruler as you did with the glass marble, is the velocity of this steel marble the same as the velocity of the marble at the bottom of the ruler? Is the distance the cup moves proportional to the mass of the steel and glass marbles?



Rolling a marble down a ruler into a foam cup.

Note:**PhET Explorations: The Ramp**

Explore forces, energy and work as you push household objects up and down a ramp. Lower and raise the ramp to see how the angle of inclination affects the parallel forces acting on the file cabinet. Graphs show forces, energy and work.

[The
Ramp
p](#)

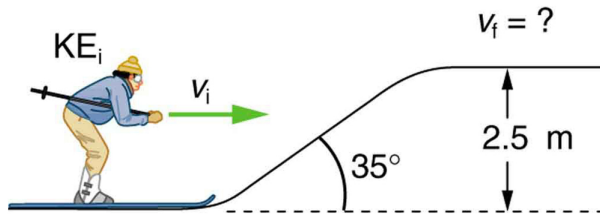
Section Summary

- A nonconservative force is one for which work depends on the path.
- Friction is an example of a nonconservative force that changes mechanical energy into thermal energy.
- Work W_{nc} done by a nonconservative force changes the mechanical energy of a system. In equation form, $W_{nc} = \Delta KE + \Delta PE$ or, equivalently, $KE_i + PE_i + W_{nc} = KE_f + PE_f$.
- When both conservative and nonconservative forces act, energy conservation can be applied and used to calculate motion in terms of the known potential energies of the conservative forces and the work done by nonconservative forces, instead of finding the net work from the net force, or having to directly apply Newton's laws.

Problems & Exercises**Exercise:**

Problem:

A 60.0-kg skier with an initial speed of 12.0 m/s coasts up a 2.50-m-high rise as shown in [\[link\]](#). Find her final speed at the top, given that the coefficient of friction between her skis and the snow is 0.0800. (Hint: Find the distance traveled up the incline assuming a straight-line path as shown in the figure.)



The skier's initial kinetic energy is partially used in coasting to the top of a rise.

Solution:

9.46 m/s

Exercise:**Problem:**

(a) How high a hill can a car coast up (engine disengaged) if work done by friction is negligible and its initial speed is 110 km/h? (b) If, in actuality, a 750-kg car with an initial speed of 110 km/h is observed to coast up a hill to a height 22.0 m above its starting point, how much thermal energy was generated by friction? (c) What is the average force of friction if the hill has a slope 2.5° above the horizontal?

Glossary

nonconservative force

a force whose work depends on the path followed between the given initial and final configurations

friction

the force between surfaces that opposes one sliding on the other;
friction changes mechanical energy into thermal energy

Conservation of Energy

- Explain the law of the conservation of energy.
- Describe some of the many forms of energy.
- Define efficiency of an energy conversion process as the fraction left as useful energy or work, rather than being transformed, for example, into thermal energy.

Law of Conservation of Energy

Energy, as we have noted, is conserved, making it one of the most important physical quantities in nature. The **law of conservation of energy** can be stated as follows:

Total energy is constant in any process. It may change in form or be transferred from one system to another, but the total remains the same.

We have explored some forms of energy and some ways it can be transferred from one system to another. This exploration led to the definition of two major types of energy—mechanical energy ($KE + PE$) and energy transferred via work done by nonconservative forces (W_{nc}). But energy takes *many* other forms, manifesting itself in *many* different ways, and we need to be able to deal with all of these before we can write an equation for the above general statement of the conservation of energy.

Other Forms of Energy than Mechanical Energy

At this point, we deal with all other forms of energy by lumping them into a single group called other energy (OE). Then we can state the conservation of energy in equation form as

Equation:

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f.$$

All types of energy and work can be included in this very general statement of conservation of energy. Kinetic energy is KE , work done by a conservative force is represented by PE , work done by nonconservative forces is W_{nc} , and

all other energies are included as OE. This equation applies to all previous examples; in those situations OE was constant, and so it subtracted out and was not directly considered.

Note:

Making Connections: Usefulness of the Energy Conservation Principle

The fact that energy is conserved and has many forms makes it very important. You will find that energy is discussed in many contexts, because it is involved in all processes. It will also become apparent that many situations are best understood in terms of energy and that problems are often most easily conceptualized and solved by considering energy.

When does OE play a role? One example occurs when a person eats. Food is oxidized with the release of carbon dioxide, water, and energy. Some of this chemical energy is converted to kinetic energy when the person moves, to potential energy when the person changes altitude, and to thermal energy (another form of OE).

Some of the Many Forms of Energy

What are some other forms of energy? You can probably name a number of forms of energy not yet discussed. Many of these will be covered in later chapters, but let us detail a few here. **Electrical energy** is a common form that is converted to many other forms and does work in a wide range of practical situations. Fuels, such as gasoline and food, carry **chemical energy** that can be transferred to a system through oxidation. Chemical fuel can also produce electrical energy, such as in batteries. Batteries can in turn produce light, which is a very pure form of energy. Most energy sources on Earth are in fact stored energy from the energy we receive from the Sun. We sometimes refer to this as **radiant energy**, or electromagnetic radiation, which includes visible light, infrared, and ultraviolet radiation. **Nuclear energy** comes from processes that convert measurable amounts of mass into energy. Nuclear energy is transformed into the energy of sunlight, into electrical energy in power plants, and into the energy of the heat transfer and blast in weapons.

Atoms and molecules inside all objects are in random motion. This internal mechanical energy from the random motions is called **thermal energy**, because it is related to the temperature of the object. These and all other forms of energy can be converted into one another and can do work.

[\[link\]](#) gives the amount of energy stored, used, or released from various objects and in various phenomena. The range of energies and the variety of types and situations is impressive.

Note:

Problem-Solving Strategies for Energy

You will find the following problem-solving strategies useful whenever you deal with energy. The strategies help in organizing and reinforcing energy concepts. In fact, they are used in the examples presented in this chapter. The familiar general problem-solving strategies presented earlier—involving identifying physical principles, knowns, and unknowns, checking units, and so on—continue to be relevant here.

Step 1. Determine the system of interest and identify what information is given and what quantity is to be calculated. A sketch will help.

Step 2. Examine all the forces involved and determine whether you know or are given the potential energy from the work done by the forces. Then use step 3 or step 4.

Step 3. If you know the potential energies for the forces that enter into the problem, then forces are all conservative, and you can apply conservation of mechanical energy simply in terms of potential and kinetic energy. The equation expressing conservation of energy is

Equation:

$$KE_i + PE_i = KE_f + PE_f.$$

Step 4. If you know the potential energy for only some of the forces, possibly because some of them are nonconservative and do not have a potential energy, or if there are other energies that are not easily treated in terms of force and work, then the conservation of energy law in its most general form must be used.

Equation:

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f.$$

In most problems, one or more of the terms is zero, simplifying its solution. Do not calculate W_c , the work done by conservative forces; it is already incorporated in the PE terms.

Step 5. You have already identified the types of work and energy involved (in step 2). Before solving for the unknown, *eliminate terms wherever possible* to simplify the algebra. For example, choose $h = 0$ at either the initial or final point, so that PE_g is zero there. Then solve for the unknown in the customary manner.

Step 6. *Check the answer to see if it is reasonable.* Once you have solved a problem, reexamine the forms of work and energy to see if you have set up the conservation of energy equation correctly. For example, work done against friction should be negative, potential energy at the bottom of a hill should be less than that at the top, and so on. Also check to see that the numerical value obtained is reasonable. For example, the final speed of a skateboarder who coasts down a 3-m-high ramp could reasonably be 20 km/h, but *not* 80 km/h.

Transformation of Energy

The transformation of energy from one form into others is happening all the time. The chemical energy in food is converted into thermal energy through metabolism; light energy is converted into chemical energy through photosynthesis. In a larger example, the chemical energy contained in coal is converted into thermal energy as it burns to turn water into steam in a boiler. This thermal energy in the steam in turn is converted to mechanical energy as it spins a turbine, which is connected to a generator to produce electrical energy. (In all of these examples, not all of the initial energy is converted into the forms mentioned. This important point is discussed later in this section.)

Another example of energy conversion occurs in a solar cell. Sunlight impinging on a solar cell (see [\[link\]](#)) produces electricity, which in turn can be used to run an electric motor. Energy is converted from the primary source of solar energy into electrical energy and then into mechanical energy.



Solar energy is converted into electrical energy by solar cells, which is used to run a motor in this solar-power aircraft. (credit: NASA)

Object/phenomenon	Energy in joules
Big Bang	10^{68}
Energy released in a supernova	10^{44}
Fusion of all the hydrogen in Earth's oceans	10^{34}
Annual world energy use	4×10^{20}

Object/phenomenon	Energy in joules
Large fusion bomb (9 megaton)	3.8×10^{16}
1 kg hydrogen (fusion to helium)	6.4×10^{14}
1 kg uranium (nuclear fission)	8.0×10^{13}
Hiroshima-size fission bomb (10 kiloton)	4.2×10^{13}
90,000-ton aircraft carrier at 30 knots	1.1×10^{10}
1 barrel crude oil	5.9×10^9
1 ton TNT	4.2×10^9
1 gallon of gasoline	1.2×10^8
Daily home electricity use (developed countries)	7×10^7
Daily adult food intake (recommended)	1.2×10^7

Object/phenomenon	Energy in joules
1000-kg car at 90 km/h	3.1×10^5
1 g fat (9.3 kcal)	3.9×10^4
ATP hydrolysis reaction	3.2×10^4
1 g carbohydrate (4.1 kcal)	1.7×10^4
1 g protein (4.1 kcal)	1.7×10^4
Tennis ball at 100 km/h	22
Mosquito (10^{-2} g at 0.5 m/s)	1.3×10^{-6}
Single electron in a TV tube beam	4.0×10^{-15}
Energy to break one DNA strand	10^{-19}

Energy of Various Objects and Phenomena

Efficiency

Even though energy is conserved in an energy conversion process, the output of *useful energy* or work will be less than the energy input. The **efficiency** Eff of an energy conversion process is defined as

Equation:

$$\text{Efficiency}(\text{Eff}) = \frac{\text{useful energy or work output}}{\text{total energy input}} = \frac{W_{\text{out}}}{E_{\text{in}}}.$$

[\[link\]](#) lists some efficiencies of mechanical devices and human activities. In a coal-fired power plant, for example, about 40% of the chemical energy in the coal becomes useful electrical energy. The other 60% transforms into other (perhaps less useful) energy forms, such as thermal energy, which is then released to the environment through combustion gases and cooling towers.

Activity/device	Efficiency (%) [footnote] Representative values
Cycling and climbing	20
Swimming, surface	2
Swimming, submerged	4
Shoveling	3
Weightlifting	9
Steam engine	17
Gasoline engine	30

Activity/device	Efficiency (%) ^[footnote] Representative values
Diesel engine	35
Nuclear power plant	35
Coal power plant	42
Electric motor	98
Compact fluorescent light	20
Gas heater (residential)	90
Solar cell	10

Efficiency of the Human Body and Mechanical Devices

Note:

PhET Explorations: Masses and Springs

A realistic mass and spring laboratory. Hang masses from springs and adjust the spring stiffness and damping. You can even slow time. Transport the lab to different planets. A chart shows the kinetic, potential, and thermal energies for each spring.

https://phet.colorado.edu/sims/mass-spring-lab/mass-spring-lab_en.html

Section Summary

- The law of conservation of energy states that the total energy is constant in any process. Energy may change in form or be transferred from one system to another, but the total remains the same.
- When all forms of energy are considered, conservation of energy is written in equation form as

$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f$, where OE is all **other forms of energy** besides mechanical energy.

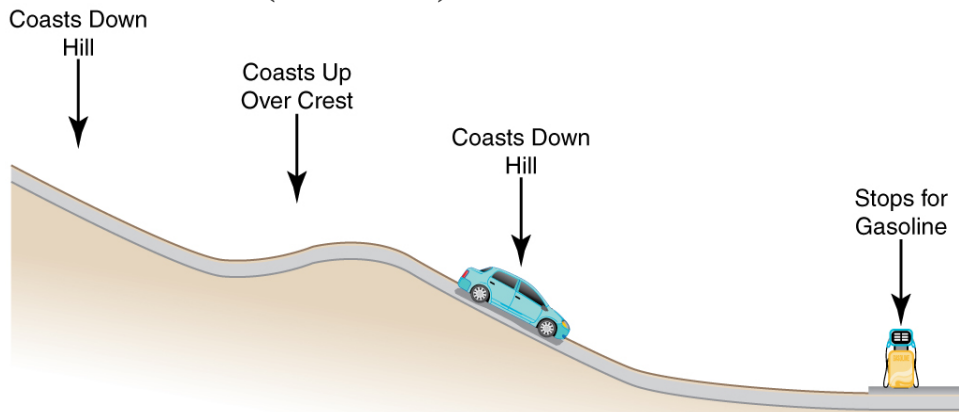
- Commonly encountered forms of energy include electric energy, chemical energy, radiant energy, nuclear energy, and thermal energy.
- Energy is often utilized to do work, but it is not possible to convert all the energy of a system to work.
- The efficiency Eff of a machine or human is defined to be $Eff = \frac{W_{out}}{E_{in}}$, where W_{out} is useful work output and E_{in} is the energy consumed.

Conceptual Questions

Exercise:

Problem:

Consider the following scenario. A car for which friction is *not* negligible accelerates from rest down a hill, running out of gasoline after a short distance. The driver lets the car coast farther down the hill, then up and over a small crest. He then coasts down that hill into a gas station, where he brakes to a stop and fills the tank with gasoline. Identify the forms of energy the car has, and how they are changed and transferred in this series of events. (See [\[link\]](#).)



A car experiencing non-negligible friction coasts down a hill, over a small crest, then downhill again, and comes to a stop at a gas station.

Exercise:**Problem:**

Describe the energy transfers and transformations for a javelin, starting from the point at which an athlete picks up the javelin and ending when the javelin is stuck into the ground after being thrown.

Exercise:**Problem:**

Do devices with efficiencies of less than one violate the law of conservation of energy? Explain.

Exercise:**Problem:**

List four different forms or types of energy. Give one example of a conversion from each of these forms to another form.

Exercise:

Problem: List the energy conversions that occur when riding a bicycle.

Problems & Exercises**Exercise:****Problem:**

Using values from [\[link\]](#), how many DNA molecules could be broken by the energy carried by a single electron in the beam of an old-fashioned TV tube? (These electrons were not dangerous in themselves, but they did create dangerous x rays. Later model tube TVs had shielding that absorbed x rays before they escaped and exposed viewers.)

Solution:

4×10^4 molecules

Exercise:**Problem:**

Using energy considerations and assuming negligible air resistance, show that a rock thrown from a bridge 20.0 m above water with an initial speed of 15.0 m/s strikes the water with a speed of 24.8 m/s independent of the direction thrown.

Solution:

Equating ΔPE_g and ΔKE , we obtain

$$v = \sqrt{2gh + v_0^2} = \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m}) + (15.0 \text{ m/s})^2} = 24.8 \text{ m/s}$$

Exercise:**Problem:**

If the energy in fusion bombs were used to supply the energy needs of the world, how many of the 9-megaton variety would be needed for a year's supply of energy (using data from [\[link\]](#))? This is not as far-fetched as it may sound—there are thousands of nuclear bombs, and their energy can be trapped in underground explosions and converted to electricity, as natural geothermal energy is.

Exercise:**Problem:**

(a) Use of hydrogen fusion to supply energy is a dream that may be realized in the next century. Fusion would be a relatively clean and almost limitless supply of energy, as can be seen from [\[link\]](#). To illustrate this, calculate how many years the present energy needs of the world could be supplied by one millionth of the oceans' hydrogen fusion energy. (b) How does this time compare with historically significant events, such as the duration of stable economic systems?

Solution:

(a) 25×10^6 years

(b) This is much, much longer than human time scales.

Glossary

law of conservation of energy

the general law that total energy is constant in any process; energy may change in form or be transferred from one system to another, but the total remains the same

electrical energy

the energy carried by a flow of charge

chemical energy

the energy in a substance stored in the bonds between atoms and molecules that can be released in a chemical reaction

radiant energy

the energy carried by electromagnetic waves

nuclear energy

energy released by changes within atomic nuclei, such as the fusion of two light nuclei or the fission of a heavy nucleus

thermal energy

the energy within an object due to the random motion of its atoms and molecules that accounts for the object's temperature

efficiency

a measure of the effectiveness of the input of energy to do work; useful energy or work divided by the total input of energy

Power

- Calculate power by calculating changes in energy over time.
- Examine power consumption and calculations of the cost of energy consumed.

What is Power?

Power—the word conjures up many images: a professional football player muscling aside his opponent, a dragster roaring away from the starting line, a volcano blowing its lava into the atmosphere, or a rocket blasting off, as in [\[link\]](#).



This powerful rocket on the Space Shuttle *Endeavor* did work and consumed energy at a very high rate. (credit: NASA)

These images of power have in common the rapid performance of work, consistent with the scientific definition of **power** (P) as the rate at which work is done.

Note:**Power**

Power is the rate at which work is done.

Equation:

$$P = \frac{W}{t}$$

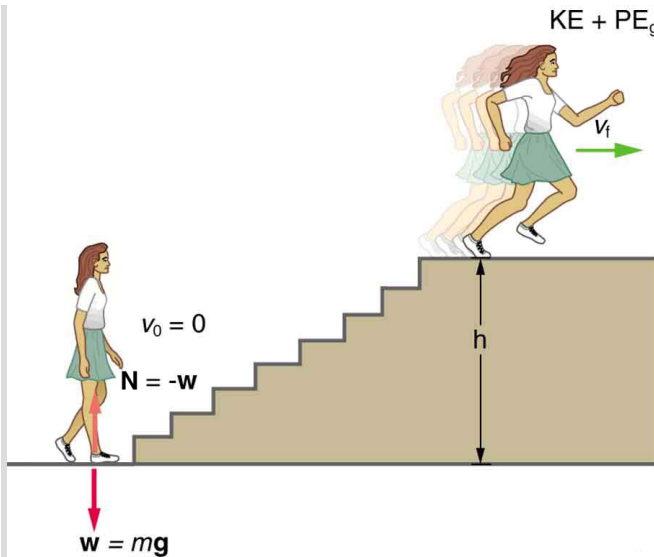
The SI unit for power is the **watt** (W), where 1 watt equals 1 joule/second ($1 \text{ W} = 1 \text{ J/s}$).

Because work is energy transfer, power is also the rate at which energy is expended. A 60-W light bulb, for example, expends 60 J of energy per second. Great power means a large amount of work or energy developed in a short time. For example, when a powerful car accelerates rapidly, it does a large amount of work and consumes a large amount of fuel in a short time.

Calculating Power from Energy

Example:**Calculating the Power to Climb Stairs**

What is the power output for a 60.0-kg woman who runs up a 3.00 m high flight of stairs in 3.50 s, starting from rest but having a final speed of 2.00 m/s? (See [\[link\]](#).)



When this woman runs upstairs starting from rest, she converts the chemical energy originally from food into kinetic energy and gravitational potential energy. Her power output depends on how fast she does this.

Strategy and Concept

The work going into mechanical energy is $W = KE + PE$. At the bottom of the stairs, we take both KE and PE_g as initially zero; thus,

$W = KE_f + PE_g = \frac{1}{2}mv_f^2 + mgh$, where h is the vertical height of the stairs. Because all terms are given, we can calculate W and then divide it by time to get power.

Solution

Substituting the expression for W into the definition of power given in the previous equation, $P = W/t$ yields

Equation:

$$P = \frac{W}{t} = \frac{\frac{1}{2}mv_f^2 + mgh}{t}.$$

Entering known values yields

Equation:

$$\begin{aligned} P &= \frac{0.5(60.0 \text{ kg})(2.00 \text{ m/s})^2 + (60.0 \text{ kg})(9.80 \text{ m/s}^2)(3.00 \text{ m})}{3.50 \text{ s}} \\ &= \frac{120 \text{ J} + 1764 \text{ J}}{3.50 \text{ s}} \\ &= 538 \text{ W}. \end{aligned}$$

Discussion

The woman does 1764 J of work to move up the stairs compared with only 120 J to increase her kinetic energy; thus, most of her power output is required for climbing rather than accelerating.

It is impressive that this woman's useful power output is slightly less than 1 **horsepower** (1 hp = 746 W)! People can generate more than a horsepower with their leg muscles for short periods of time by rapidly converting available blood sugar and oxygen into work output. (A horse can put out 1 hp for hours on end.) Once oxygen is depleted, power output decreases and the person begins to breathe rapidly to obtain oxygen to metabolize more food—this is known as the *aerobic* stage of exercise. If the woman climbed the stairs slowly, then her power output would be much less, although the amount of work done would be the same.

Note:**Making Connections: Take-Home Investigation—Measure Your Power Rating**

Determine your own power rating by measuring the time it takes you to climb a flight of stairs. We will ignore the gain in kinetic energy, as the above example showed that it was a small portion of the energy gain. Don't expect that your output will be more than about 0.5 hp.

Examples of Power

Examples of power are limited only by the imagination, because there are as many types as there are forms of work and energy. (See [\[link\]](#) for some examples.) Sunlight reaching Earth's surface carries a maximum power of about 1.3 kilowatts per square meter (kW/m^2). A tiny fraction of this is retained by Earth over the long term. Our consumption rate of fossil fuels is far greater than the rate at which they are stored, so it is inevitable that they will be depleted. Power implies that energy is transferred, perhaps changing form. It is never possible to change one form completely into another without losing some of it as thermal energy. For example, a 60-W incandescent bulb converts only 5 W of electrical power to light, with 55 W dissipating into thermal energy. Furthermore, the typical electric power plant converts only 35 to 40% of its fuel into electricity. The remainder becomes a huge amount of thermal energy that must be dispersed as heat transfer, as rapidly as it is created. A coal-fired power plant may produce 1000 megawatts; 1 megawatt (MW) is 10^6 W of electric power. But the power plant consumes chemical energy at a rate of about 2500 MW, creating heat transfer to the surroundings at a rate of 1500 MW. (See [\[link\]](#).)



Tremendous amounts of electric power are generated by coal-fired power plants such as this one in China, but an even larger amount of power goes into heat transfer to the surroundings.

The large cooling towers here are needed to transfer heat as rapidly as it is produced. The transfer of heat is not unique to coal plants but is an unavoidable consequence of generating electric power from any fuel—nuclear, coal, oil, natural gas, or the like. (credit: Kleinolive, Wikimedia Commons)

Object or Phenomenon	Power in Watts
Supernova (at peak)	5×10^{37}
Milky Way galaxy	10^{37}
Crab Nebula pulsar	10^{28}
The Sun	4×10^{26}

Object or Phenomenon	Power in Watts
Volcanic eruption (maximum)	4×10^{15}
Lightning bolt	2×10^{12}
Nuclear power plant (total electric and heat transfer)	3×10^9
Aircraft carrier (total useful and heat transfer)	10^8
Dragster (total useful and heat transfer)	2×10^6
Car (total useful and heat transfer)	8×10^4
Football player (total useful and heat transfer)	5×10^3
Clothes dryer	4×10^3
Person at rest (all heat transfer)	100

Object or Phenomenon	Power in Watts
Typical incandescent light bulb (total useful and heat transfer)	60
Heart, person at rest (total useful and heat transfer)	8
Electric clock	3
Pocket calculator	10^{-3}

Power Output or Consumption

Power and Energy Consumption

We usually have to pay for the energy we use. It is interesting and easy to estimate the cost of energy for an electrical appliance if its power consumption rate and time used are known. The higher the power consumption rate and the longer the appliance is used, the greater the cost of that appliance. The power consumption rate is $P = W/t = E/t$, where E is the energy supplied by the electricity company. So the energy consumed over a time t is

Equation:

$$E = Pt.$$

Electricity bills state the energy used in units of **kilowatt-hours** ($\text{kW} \cdot \text{h}$), which is the product of power in kilowatts and time in hours. This unit is convenient because electrical power consumption at the kilowatt level for hours at a time is typical.

Example:**Calculating Energy Costs**

What is the cost of running a 0.200-kW computer 6.00 h per day for 30.0 d if the cost of electricity is \$0.120 per kW · h?

Strategy

Cost is based on energy consumed; thus, we must find E from $E = Pt$ and then calculate the cost. Because electrical energy is expressed in kW · h, at the start of a problem such as this it is convenient to convert the units into kW and hours.

Solution

The energy consumed in kW · h is

Equation:

$$\begin{aligned} E &= Pt = (0.200 \text{ kW})(6.00 \text{ h/d})(30.0 \text{ d}) \\ &= 36.0 \text{ kW} \cdot \text{h}, \end{aligned}$$

and the cost is simply given by

Equation:

$$\text{cost} = (36.0 \text{ kW} \cdot \text{h})(\$0.120 \text{ per kW} \cdot \text{h}) = \$4.32 \text{ per month.}$$

Discussion

The cost of using the computer in this example is neither exorbitant nor negligible. It is clear that the cost is a combination of power and time. When both are high, such as for an air conditioner in the summer, the cost is high.

The motivation to save energy has become more compelling with its ever-increasing price. Armed with the knowledge that energy consumed is the product of power and time, you can estimate costs for yourself and make the necessary value judgments about where to save energy. Either power or time must be reduced. It is most cost-effective to limit the use of high-power devices that normally operate for long periods of time, such as water heaters and air conditioners. This would not include relatively high power devices like toasters, because they are on only a few minutes per day. It would also not include electric clocks, in spite of their 24-hour-per-day

usage, because they are very low power devices. It is sometimes possible to use devices that have greater efficiencies—that is, devices that consume less power to accomplish the same task. One example is the compact fluorescent light bulb, which produces over four times more light per watt of power consumed than its incandescent cousin.

Modern civilization depends on energy, but current levels of energy consumption and production are not sustainable. The likelihood of a link between global warming and fossil fuel use (with its concomitant production of carbon dioxide), has made reduction in energy use as well as a shift to non-fossil fuels of the utmost importance. Even though energy in an isolated system is a conserved quantity, the final result of most energy transformations is waste heat transfer to the environment, which is no longer useful for doing work. As we will discuss in more detail in [Thermodynamics](#), the potential for energy to produce useful work has been “degraded” in the energy transformation.

Section Summary

- Power is the rate at which work is done, or in equation form, for the average power P for work W done over a time t , $P = W/t$.
- The SI unit for power is the watt (W), where $1 \text{ W} = 1 \text{ J/s}$.
- The power of many devices such as electric motors is also often expressed in horsepower (hp), where $1 \text{ hp} = 746 \text{ W}$.

Conceptual Questions

Exercise:

Problem:

Most electrical appliances are rated in watts. Does this rating depend on how long the appliance is on? (When off, it is a zero-watt device.) Explain in terms of the definition of power.

Exercise:

Problem:

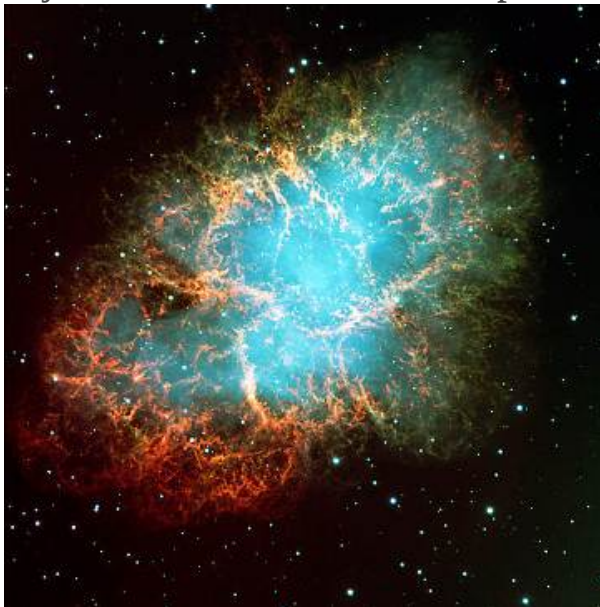
Explain, in terms of the definition of power, why energy consumption is sometimes listed in kilowatt-hours rather than joules. What is the relationship between these two energy units?

Exercise:**Problem:**

A spark of static electricity, such as that you might receive from a doorknob on a cold dry day, may carry a few hundred watts of power. Explain why you are not injured by such a spark.

Problems & Exercises**Exercise:****Problem:**

The Crab Nebula (see [\[link\]](#)) pulsar is the remnant of a supernova that occurred in A.D. 1054. Using data from [\[link\]](#), calculate the approximate factor by which the power output of this astronomical object has declined since its explosion.



Crab Nebula (credit: ESO, via
Wikimedia Commons)

Solution:
Equation:

$$2 \times 10^{-10}$$

Exercise:

Problem:

Suppose a star 1000 times brighter than our Sun (that is, emitting 1000 times the power) suddenly goes supernova. Using data from [\[link\]](#): (a) By what factor does its power output increase? (b) How many times brighter than our entire Milky Way galaxy is the supernova? (c) Based on your answers, discuss whether it should be possible to observe supernovas in distant galaxies. Note that there are on the order of 10^{11} observable galaxies, the average brightness of which is somewhat less than our own galaxy.

Exercise:

Problem:

A person in good physical condition can put out 100 W of useful power for several hours at a stretch, perhaps by pedaling a mechanism that drives an electric generator. Neglecting any problems of generator efficiency and practical considerations such as resting time: (a) How many people would it take to run a 4.00-kW electric clothes dryer? (b) How many people would it take to replace a large electric power plant that generates 800 MW?

Solution:

(a) 40

(b) 8 million

Exercise:

Problem:

What is the cost of operating a 3.00-W electric clock for a year if the cost of electricity is \$0.0900 per $\text{kW} \cdot \text{h}$?

Exercise:

Problem:

A large household air conditioner may consume 15.0 kW of power. What is the cost of operating this air conditioner 3.00 h per day for 30.0 d if the cost of electricity is \$0.110 per $\text{kW} \cdot \text{h}$?

Solution:

\$149

Exercise:

Problem:

(a) What is the average power consumption in watts of an appliance that uses 5.00 $\text{kW} \cdot \text{h}$ of energy per day? (b) How many joules of energy does this appliance consume in a year?

Exercise:

Problem:

(a) What is the average useful power output of a person who does $6.00 \times 10^6 \text{ J}$ of useful work in 8.00 h? (b) Working at this rate, how long will it take this person to lift 2000 kg of bricks 1.50 m to a platform? (Work done to lift his body can be omitted because it is not considered useful output here.)

Solution:

(a) 208 W

(b) 141 s

Exercise:

Problem:

A 500-kg dragster accelerates from rest to a final speed of 110 m/s in 400 m (about a quarter of a mile) and encounters an average frictional force of 1200 N. What is its average power output in watts and horsepower if this takes 7.30 s?

Exercise:

Problem:

(a) How long will it take an 850-kg car with a useful power output of 40.0 hp (1 hp = 746 W) to reach a speed of 15.0 m/s, neglecting friction? (b) How long will this acceleration take if the car also climbs a 3.00-m-high hill in the process?

Solution:

(a) 3.20 s

(b) 4.04 s

Exercise:

Problem:

(a) Find the useful power output of an elevator motor that lifts a 2500-kg load a height of 35.0 m in 12.0 s, if it also increases the speed from rest to 4.00 m/s. Note that the total mass of the counterbalanced system is 10,000 kg—so that only 2500 kg is raised in height, but the full 10,000 kg is accelerated. (b) What does it cost, if electricity is \$0.0900 per kW · h?

Exercise:

Problem:

(a) What is the available energy content, in joules, of a battery that operates a 2.00-W electric clock for 18 months? (b) How long can a battery that can supply 8.00×10^4 J run a pocket calculator that consumes energy at the rate of 1.00×10^{-3} W?

Solution:

(a) 9.46×10^7 J

(b) 2.54 y

Exercise:**Problem:**

(a) How long would it take a 1.50×10^5 -kg airplane with engines that produce 100 MW of power to reach a speed of 250 m/s and an altitude of 12.0 km if air resistance were negligible? (b) If it actually takes 900 s, what is the power? (c) Given this power, what is the average force of air resistance if the airplane takes 1200 s? (Hint: You must find the distance the plane travels in 1200 s assuming constant acceleration.)

Exercise:**Problem:**

Calculate the power output needed for a 950-kg car to climb a 2.00° slope at a constant 30.0 m/s while encountering wind resistance and friction totaling 600 N. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Energy](#).

Solution:

Identify knowns: $m = 950$ kg, slope angle $\theta = 2.00^\circ$, $v = 30.0$ m/s, $f = 600$ N

Identify unknowns: power P of the car, force F that car applies to road

Solve for unknown:

$$P = \frac{W}{t} = \frac{Fd}{t} = F\left(\frac{d}{t}\right) = Fv,$$

where F is parallel to the incline and must oppose the resistive forces and the force of gravity:

$$F = f + w = 600 \text{ N} + mg \sin \theta$$

Insert this into the expression for power and solve:

$$\begin{aligned} P &= (f + mg \sin \theta)v \\ &= \left[600 \text{ N} + (950 \text{ kg}) \left(9.80 \text{ m/s}^2 \right) \sin 2^\circ \right] (30.0 \text{ m/s}) \\ &= 2.77 \times 10^4 \text{ W} \end{aligned}$$

About 28 kW (or about 37 hp) is reasonable for a car to climb a gentle incline.

Exercise:

Problem:

(a) Calculate the power per square meter reaching Earth's upper atmosphere from the Sun. (Take the power output of the Sun to be $4.00 \times 10^{26} \text{ W}$.) (b) Part of this is absorbed and reflected by the atmosphere, so that a maximum of 1.30 kW/m^2 reaches Earth's surface. Calculate the area in km^2 of solar energy collectors needed to replace an electric power plant that generates 750 MW if the collectors convert an average of 2.00% of the maximum power into electricity. (This small conversion efficiency is due to the devices themselves, and the fact that the sun is directly overhead only briefly.) With the same assumptions, what area would be needed to meet the United States' energy needs ($1.05 \times 10^{20} \text{ J}$)? Australia's energy needs ($5.4 \times 10^{18} \text{ J}$)? China's energy needs ($6.3 \times 10^{19} \text{ J}$)? (These energy consumption values are from 2006.)

Glossary

power

the rate at which work is done

watt

(W) SI unit of power, with $1 \text{ W} = 1 \text{ J/s}$

horsepower

an older non-SI unit of power, with $1 \text{ hp} = 746 \text{ W}$

kilowatt-hour

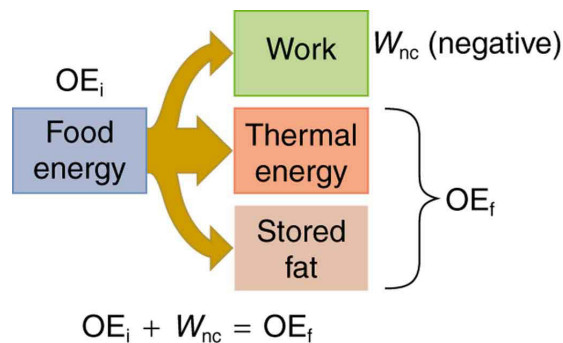
(kW · h) unit used primarily for electrical energy provided by electric utility companies

Work, Energy, and Power in Humans

- Explain the human body's consumption of energy when at rest vs. when engaged in activities that do useful work.
- Calculate the conversion of chemical energy in food into useful work.

Energy Conversion in Humans

Our own bodies, like all living organisms, are energy conversion machines. Conservation of energy implies that the chemical energy stored in food is converted into work, thermal energy, and/or stored as chemical energy in fatty tissue. (See [\[link\]](#).) The fraction going into each form depends both on how much we eat and on our level of physical activity. If we eat more than is needed to do work and stay warm, the remainder goes into body fat.



Energy consumed by humans is converted to work, thermal energy, and stored fat. By far the largest fraction goes to thermal energy, although the fraction varies depending on the type of physical activity.

Power Consumed at Rest

The *rate* at which the body uses food energy to sustain life and to do different activities is called the **metabolic rate**. The total energy conversion rate of a person *at rest* is called the **basal metabolic rate** (BMR) and is divided among various systems in the body, as shown in [\[link\]](#). The largest fraction goes to the liver and spleen, with the brain coming next. Of course, during vigorous exercise, the energy consumption of the skeletal muscles and heart increase markedly. About 75% of the calories burned in a day go into these basic functions. The BMR is a function of age, gender, total body weight, and amount of muscle mass (which burns more calories than body fat). Athletes have a greater BMR due to this last factor.

Organ	Power consumed at rest (W)	Oxygen consumption (mL/min)	Percent of BMR
Liver & spleen	23	67	27
Brain	16	47	19
Skeletal muscle	15	45	18
Kidney	9	26	10
Heart	6	17	7
Other	16	48	19
Totals	85 W	250 mL/min	100%

Basal Metabolic Rates (BMR)

Energy consumption is directly proportional to oxygen consumption because the digestive process is basically one of oxidizing food. We can measure the energy people use during various activities by measuring their oxygen use. (See [\[link\]](#).) Approximately 20 kJ of energy are produced for each liter of oxygen consumed, independent of the type of food. [\[link\]](#) shows energy and oxygen consumption rates (power expended) for a variety of activities.

Power of Doing Useful Work

Work done by a person is sometimes called **useful work**, which is *work done on the outside world*, such as lifting weights. Useful work requires a force exerted through a distance on the outside world, and so it excludes internal work, such as that done by the heart when pumping blood. Useful work does include that done in climbing stairs or accelerating to a full run, because these are accomplished by exerting forces on the outside world. Forces exerted by the body are nonconservative, so that they can change the mechanical energy ($KE + PE$) of the system worked upon, and this is often the goal. A baseball player throwing a ball, for example, increases both the ball's kinetic and potential energy.

If a person needs more energy than they consume, such as when doing vigorous work, the body must draw upon the chemical energy stored in fat. So exercise can be helpful in losing fat. However, the amount of exercise needed to produce a loss in fat, or to burn off extra calories consumed that day, can be large, as [\[link\]](#) illustrates.

Example:

Calculating Weight Loss from Exercising

If a person who normally requires an average of 12,000 kJ (3000 kcal) of food energy per day consumes 13,000 kJ per day, he will steadily gain weight. How much bicycling per day is required to work off this extra 1000 kJ?

Solution

[\[link\]](#) states that 400 W are used when cycling at a moderate speed. The time required to work off 1000 kJ at this rate is then

Equation:

$$\text{Time} = \frac{\text{energy}}{\left(\frac{\text{energy}}{\text{time}}\right)} = \frac{1000 \text{ kJ}}{400 \text{ W}} = 2500 \text{ s} = 42 \text{ min.}$$

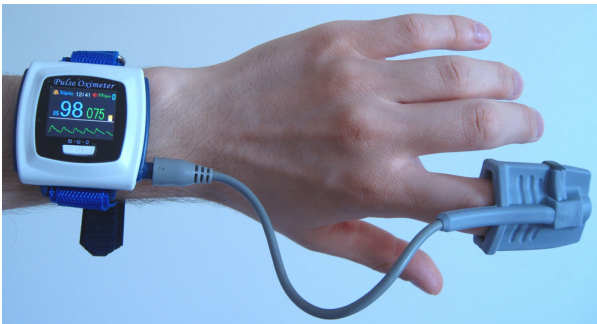
Discussion

If this person uses more energy than he or she consumes, the person's body will obtain the needed energy by metabolizing body fat. If the person uses 13,000 kJ but consumes only 12,000 kJ, then the amount of fat loss will be

Equation:

$$\text{Fat loss} = (1000 \text{ kJ}) \left(\frac{1.0 \text{ g fat}}{39 \text{ kJ}} \right) = 26 \text{ g,}$$

assuming the energy content of fat to be 39 kJ/g.



A pulse oxymeter is an apparatus that measures the amount of oxygen in blood.

Oxymeters can be used to determine a person's metabolic rate, which is the rate at which food energy is converted to another form. Such

measurements can indicate the level of athletic conditioning as well as certain medical problems. (credit: UusiAjaja, Wikimedia Commons)

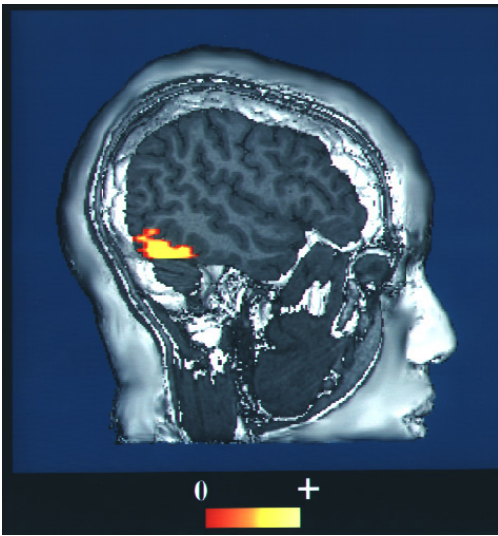
Activity	Energy consumption in watts	Oxygen consumption in liters O₂/min
Sleeping	83	0.24
Sitting at rest	120	0.34
Standing relaxed	125	0.36
Sitting in class	210	0.60
Walking (5 km/h)	280	0.80
Cycling (13–18 km/h)	400	1.14
Shivering	425	1.21
Playing tennis	440	1.26

Activity	Energy consumption in watts	Oxygen consumption in liters O₂/min
Swimming breaststroke	475	1.36
Ice skating (14.5 km/h)	545	1.56
Climbing stairs (116/min)	685	1.96
Cycling (21 km/h)	700	2.00
Running cross-country	740	2.12
Playing basketball	800	2.28
Cycling, professional racer	1855	5.30
Sprinting	2415	6.90

Energy and Oxygen Consumption Rates[\[footnote\]](#) (Power) for an average 76-kg male

All bodily functions, from thinking to lifting weights, require energy. (See [\[link\]](#).) The many small muscle actions accompanying all quiet activity, from sleeping to head scratching, ultimately become thermal energy, as do less visible muscle actions by the heart, lungs, and digestive tract. Shivering, in fact, is an involuntary response to low body temperature that pits muscles against one another to produce thermal energy in the body (and

do no work). The kidneys and liver consume a surprising amount of energy, but the biggest surprise of all is that a full 25% of all energy consumed by the body is used to maintain electrical potentials in all living cells. (Nerve cells use this electrical potential in nerve impulses.) This bioelectrical energy ultimately becomes mostly thermal energy, but some is utilized to power chemical processes such as in the kidneys and liver, and in fat production.



This fMRI scan shows an increased level of energy consumption in the vision center of the brain. Here, the patient was being asked to recognize faces.
(credit: NIH via Wikimedia Commons)

Section Summary

- The human body converts energy stored in food into work, thermal energy, and/or chemical energy that is stored in fatty tissue.
- The *rate* at which the body uses food energy to sustain life and to do different activities is called the metabolic rate, and the corresponding rate when at rest is called the basal metabolic rate (BMR)
- The energy included in the basal metabolic rate is divided among various systems in the body, with the largest fraction going to the liver and spleen, and the brain coming next.
- About 75% of food calories are used to sustain basic body functions included in the basal metabolic rate.
- The energy consumption of people during various activities can be determined by measuring their oxygen use, because the digestive process is basically one of oxidizing food.

Conceptual Questions

Exercise:

Problem:

Explain why it is easier to climb a mountain on a zigzag path rather than one straight up the side. Is your increase in gravitational potential energy the same in both cases? Is your energy consumption the same in both?

Exercise:

Problem:

Do you do work on the outside world when you rub your hands together to warm them? What is the efficiency of this activity?

Exercise:

Problem:

Shivering is an involuntary response to lowered body temperature. What is the efficiency of the body when shivering, and is this a desirable value?

Exercise:

Problem:

Discuss the relative effectiveness of dieting and exercise in losing weight, noting that most athletic activities consume food energy at a rate of 400 to 500 W, while a single cup of yogurt can contain 1360 kJ (325 kcal). Specifically, is it likely that exercise alone will be sufficient to lose weight? You may wish to consider that regular exercise may increase the metabolic rate, whereas protracted dieting may reduce it.

Problems & Exercises**Exercise:****Problem:**

(a) How long can you rapidly climb stairs (116/min) on the 93.0 kcal of energy in a 10.0-g pat of butter? (b) How many flights is this if each flight has 16 stairs?

Solution:

(a) 9.5 min

(b) 69 flights of stairs

Exercise:**Problem:**

(a) What is the power output in watts and horsepower of a 70.0-kg sprinter who accelerates from rest to 10.0 m/s in 3.00 s? (b) Considering the amount of power generated, do you think a well-trained athlete could do this repetitively for long periods of time?

Exercise:

Problem:

Calculate the power output in watts and horsepower of a shot-putter who takes 1.20 s to accelerate the 7.27-kg shot from rest to 14.0 m/s, while raising it 0.800 m. (Do not include the power produced to accelerate his body.)



Shot putter at the
Dornoch Highland
Gathering in 2007.
(credit: John Haslam,
Flickr)

Solution:

641 W, 0.860 hp

Exercise:**Problem:**

(a) What is the efficiency of an out-of-condition professor who does 2.10×10^5 J of useful work while metabolizing 500 kcal of food energy? (b) How many food calories would a well-conditioned athlete metabolize in doing the same work with an efficiency of 20%?

Exercise:

Problem:

Energy that is not utilized for work or heat transfer is converted to the chemical energy of body fat containing about 39 kJ/g. How many grams of fat will you gain if you eat 10,000 kJ (about 2500 kcal) one day and do nothing but sit relaxed for 16.0 h and sleep for the other 8.00 h? Use data from [\[link\]](#) for the energy consumption rates of these activities.

Solution:

31 g

Exercise:**Problem:**

Using data from [\[link\]](#), calculate the daily energy needs of a person who sleeps for 7.00 h, walks for 2.00 h, attends classes for 4.00 h, cycles for 2.00 h, sits relaxed for 3.00 h, and studies for 6.00 h. (Studying consumes energy at the same rate as sitting in class.)

Exercise:**Problem:**

What is the efficiency of a subject on a treadmill who puts out work at the rate of 100 W while consuming oxygen at the rate of 2.00 L/min? (Hint: See [\[link\]](#).)

Solution:

14.3%

Exercise:

Problem:

Shoveling snow can be extremely taxing because the arms have such a low efficiency in this activity. Suppose a person shoveling a footpath metabolizes food at the rate of 800 W. (a) What is her useful power output? (b) How long will it take her to lift 3000 kg of snow 1.20 m? (This could be the amount of heavy snow on 20 m of footpath.) (c) How much waste heat transfer in kilojoules will she generate in the process?

Exercise:**Problem:**

Very large forces are produced in joints when a person jumps from some height to the ground. (a) Calculate the magnitude of the force produced if an 80.0-kg person jumps from a 0.600-m-high ledge and lands stiffly, compressing joint material 1.50 cm as a result. (Be certain to include the weight of the person.) (b) In practice the knees bend almost involuntarily to help extend the distance over which you stop. Calculate the magnitude of the force produced if the stopping distance is 0.300 m. (c) Compare both forces with the weight of the person.

Solution:

(a) $3.21 \times 10^4 \text{ N}$

(b) $2.35 \times 10^3 \text{ N}$

(c) Ratio of net force to weight of person is 41.0 in part (a); 3.00 in part (b)

Exercise:

Problem:

Jogging on hard surfaces with insufficiently padded shoes produces large forces in the feet and legs. (a) Calculate the magnitude of the force needed to stop the downward motion of a jogger's leg, if his leg has a mass of 13.0 kg, a speed of 6.00 m/s, and stops in a distance of 1.50 cm. (Be certain to include the weight of the 75.0-kg jogger's body.) (b) Compare this force with the weight of the jogger.

Exercise:**Problem:**

(a) Calculate the energy in kJ used by a 55.0-kg woman who does 50 deep knee bends in which her center of mass is lowered and raised 0.400 m. (She does work in both directions.) You may assume her efficiency is 20%. (b) What is the average power consumption rate in watts if she does this in 3.00 min?

Solution:

(a) 108 kJ

(b) 599 W

Exercise:**Problem:**

Kanellos Kanellopoulos flew 119 km from Crete to Santorini, Greece, on April 23, 1988, in the *Daedalus 88*, an aircraft powered by a bicycle-type drive mechanism (see [\[link\]](#)). His useful power output for the 234-min trip was about 350 W. Using the efficiency for cycling from [\[link\]](#), calculate the food energy in kilojoules he metabolized during the flight.

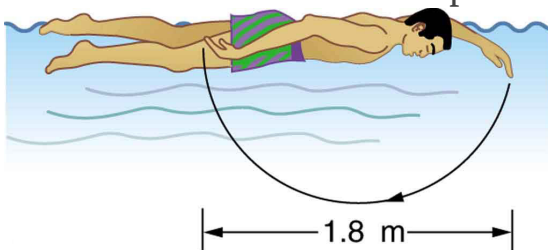


The Daedalus 88 in flight.
(credit: NASA photo by
Beasley)

Exercise:

Problem:

The swimmer shown in [\[link\]](#) exerts an average horizontal backward force of 80.0 N with his arm during each 1.80 m long stroke. (a) What is his work output in each stroke? (b) Calculate the power output of his arms if he does 120 strokes per minute.



Solution:

(a) 144 J

(b) 288 W

Exercise:

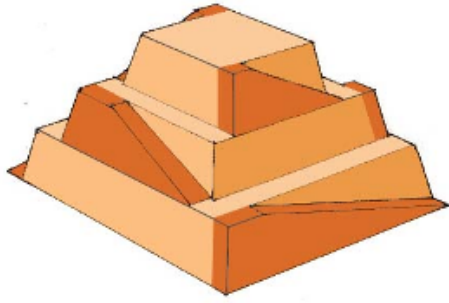
Problem:

Mountain climbers carry bottled oxygen when at very high altitudes.

(a) Assuming that a mountain climber uses oxygen at twice the rate for climbing 116 stairs per minute (because of low air temperature and winds), calculate how many liters of oxygen a climber would need for 10.0 h of climbing. (These are liters at sea level.) Note that only 40% of the inhaled oxygen is utilized; the rest is exhaled. (b) How much useful work does the climber do if he and his equipment have a mass of 90.0 kg and he gains 1000 m of altitude? (c) What is his efficiency for the 10.0-h climb?

Exercise:**Problem:**

The awe-inspiring Great Pyramid of Cheops was built more than 4500 years ago. Its square base, originally 230 m on a side, covered 13.1 acres, and it was 146 m high, with a mass of about 7×10^9 kg. (The pyramid's dimensions are slightly different today due to quarrying and some sagging.) Historians estimate that 20,000 workers spent 20 years to construct it, working 12-hour days, 330 days per year. (a) Calculate the gravitational potential energy stored in the pyramid, given its center of mass is at one-fourth its height. (b) Only a fraction of the workers lifted blocks; most were involved in support services such as building ramps (see [\[link\]](#)), bringing food and water, and hauling blocks to the site. Calculate the efficiency of the workers who did the lifting, assuming there were 1000 of them and they consumed food energy at the rate of 300 kcal/h. What does your answer imply about how much of their work went into block-lifting, versus how much work went into friction and lifting and lowering their own bodies? (c) Calculate the mass of food that had to be supplied each day, assuming that the average worker required 3600 kcal per day and that their diet was 5% protein, 60% carbohydrate, and 35% fat. (These proportions neglect the mass of bulk and nondigestible materials consumed.)



Ancient pyramids were probably constructed using ramps as simple machines.
(credit: Franck Monnier, Wikimedia Commons)

Solution:

- (a) 2.50×10^{12} J
- (b) 2.52%
- (c) 1.4×10^4 kg (14 metric tons)

Exercise:

Problem:

(a) How long can you play tennis on the 800 kJ (about 200 kcal) of energy in a candy bar? (b) Does this seem like a long time? Discuss why exercise is necessary but may not be sufficient to cause a person to lose weight.

Glossary

metabolic rate

the rate at which the body uses food energy to sustain life and to do different activities

basal metabolic rate

the total energy conversion rate of a person at rest

useful work

work done on an external system

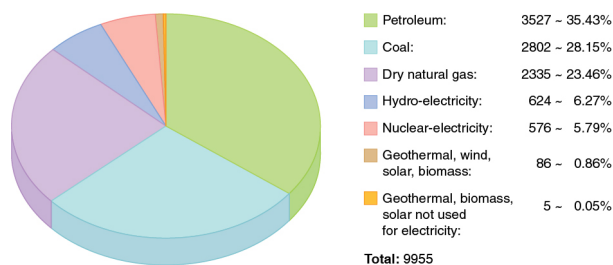
World Energy Use

- Describe the distinction between renewable and nonrenewable energy sources.
- Explain why the inevitable conversion of energy to less useful forms makes it necessary to conserve energy resources.

Energy is an important ingredient in all phases of society. We live in a very interdependent world, and access to adequate and reliable energy resources is crucial for economic growth and for maintaining the quality of our lives. But current levels of energy consumption and production are not sustainable. About 40% of the world's energy comes from oil, and much of that goes to transportation uses. Oil prices are dependent as much upon new (or foreseen) discoveries as they are upon political events and situations around the world. The U.S., with 4.5% of the world's population, consumes 24% of the world's oil production per year; 66% of that oil is imported!

Renewable and Nonrenewable Energy Sources

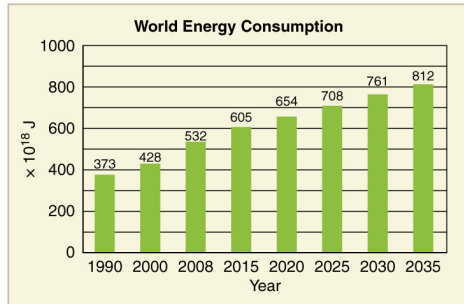
The principal energy resources used in the world are shown in [\[link\]](#). The fuel mix has changed over the years but now is dominated by oil, although natural gas and solar contributions are increasing. **Renewable forms of energy** are those sources that cannot be used up, such as water, wind, solar, and biomass. About 85% of our energy comes from nonrenewable **fossil fuels**—oil, natural gas, coal. The likelihood of a link between global warming and fossil fuel use, with its production of carbon dioxide through combustion, has made, in the eyes of many scientists, a shift to non-fossil fuels of utmost importance—but it will not be easy.



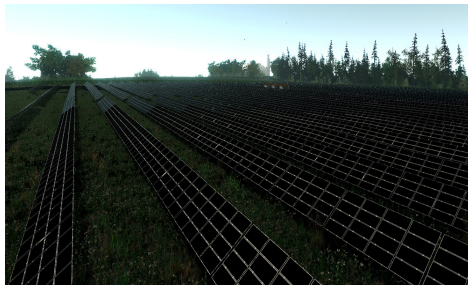
World energy consumption by source, in billions of kilowatt-hours: 2006. (credit: KVDP)

The World's Growing Energy Needs

World energy consumption continues to rise, especially in the developing countries. (See [\[link\]](#).) Global demand for energy has tripled in the past 50 years and might triple again in the next 30 years. While much of this growth will come from the rapidly booming economies of China and India, many of the developed countries, especially those in Europe, are hoping to meet their energy needs by expanding the use of renewable sources. Although presently only a small percentage, renewable energy is growing very fast, especially wind energy. For example, Germany plans to meet 20% of its electricity and 10% of its overall energy needs with renewable resources by the year 2020. (See [\[link\]](#).) Energy is a key constraint in the rapid economic growth of China and India. In 2003, China surpassed Japan as the world's second largest consumer of oil. However, over 1/3 of this is imported. Unlike most Western countries, coal dominates the commercial energy resources of China, accounting for 2/3 of its energy consumption. In 2009 China surpassed the United States as the largest generator of CO₂. In India, the main energy resources are biomass (wood and dung) and coal. Half of India's oil is imported. About 70% of India's electricity is generated by highly polluting coal. Yet there are sizeable strides being made in renewable energy. India has a rapidly growing wind energy base, and it has the largest solar cooking program in the world.



Past and projected world energy use
(source: Based on data from U.S.
Energy Information Administration,
2011)



Solar cell arrays at a power plant in
Steindorf, Germany (credit: Michael
Betke, Flickr)

[\[link\]](#) displays the 2006 commercial energy mix by country for some of the prime energy users in the world. While non-renewable sources dominate, some countries get a sizeable percentage of their electricity from renewable resources. For example, about 67% of New Zealand's electricity demand is met by hydroelectric. Only 10% of the U.S. electricity is generated by renewable resources, primarily hydroelectric. It is difficult to determine total contributions of renewable energy in some countries with a large rural population, so these percentages in this table are left blank.

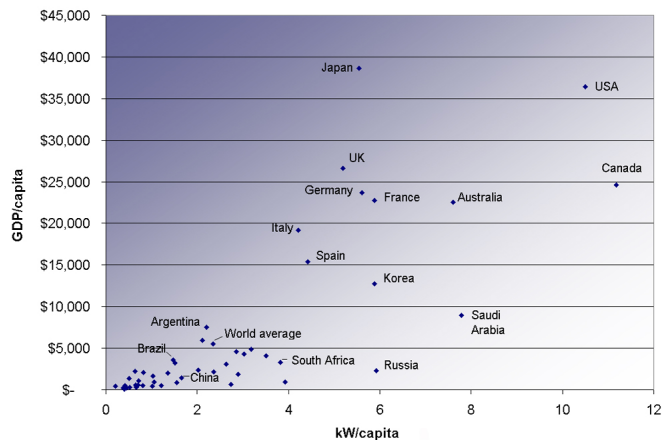
Country	Consumption, in EJ (10 ¹⁸ J)	Oil	Natural Gas	Coal	Nuclear	Hydro	Other Renewables
Australia	5.4	34%	17%	44%	0%	3%	1%

Country	Consumption, in EJ (10^{18} J)	Oil	Natural Gas	Coal	Nuclear	Hydro	Other Renewables
Brazil	9.6	48%	7%	5%	1%	35%	2%
China	63	22%	3%	69%	1%	6%	
Egypt	2.4	50%	41%	1%	0%	6%	
Germany	16	37%	24%	24%	11%	1%	3%
India	15	34%	7%	52%	1%	5%	
Indonesia	4.9	51%	26%	16%	0%	2%	3%
Japan	24	48%	14%	21%	12%	4%	1%
New Zealand	0.44	32%	26%	6%	0%	11%	19%
Russia	31	19%	53%	16%	5%	6%	
U.S.	105	40%	23%	22%	8%	3%	1%
World	432	39%	23%	24%	6%	6%	2%

Energy Consumption—Selected Countries (2006)

Energy and Economic Well-being

The last two columns in this table examine the energy and electricity use per capita. Economic well-being is dependent upon energy use, and in most countries higher standards of living, as measured by GDP (gross domestic product) per capita, are matched by higher levels of energy consumption per capita. This is borne out in [\[link\]](#). Increased efficiency of energy use will change this dependency. A global problem is balancing energy resource development against the harmful effects upon the environment in its extraction and use.



Power consumption per capita versus GDP per capita for various countries. Note the increase in energy usage with increasing GDP. (2007, credit: Frank van Mierlo, Wikimedia Commons)

Conserving Energy

As we finish this chapter on energy and work, it is relevant to draw some distinctions between two sometimes misunderstood terms in the area of energy use. As has been mentioned elsewhere, the “law of the conservation of energy” is a very useful principle in analyzing physical processes. It is a statement that cannot be proven from basic principles, but is a very good bookkeeping device, and no exceptions have ever been found. It states that the total amount of energy in an isolated system will always remain constant. Related to this principle, but remarkably different from it, is the important philosophy of energy conservation. This concept has to do with seeking to decrease the amount of energy used by an individual or group through (1) reduced activities (e.g., turning down thermostats, driving fewer kilometers) and/or (2) increasing conversion efficiencies in the performance of a particular task—such as developing and using more efficient room heaters, cars that have greater miles-per-gallon ratings, energy-efficient compact fluorescent lights, etc.

Since energy in an isolated system is not destroyed or created or generated, one might wonder why we need to be concerned about our energy resources, since energy is a conserved quantity. The problem is that the final result of most energy transformations is waste heat transfer to the environment and conversion to energy forms no longer useful for doing work. To state it in another way, the potential for energy to produce useful work has been “degraded” in the energy transformation. (This will be discussed in more detail in [Thermodynamics](#).)

Section Summary

- The relative use of different fuels to provide energy has changed over the years, but fuel use is currently dominated by oil, although natural gas and solar contributions are increasing.
- Although non-renewable sources dominate, some countries meet a sizeable percentage of their electricity needs from renewable resources.
- The United States obtains only about 10% of its energy from renewable sources, mostly hydroelectric power.
- Economic well-being is dependent upon energy use, and in most countries higher standards of living, as measured by GDP (Gross Domestic Product) per capita, are matched by higher levels of energy consumption per capita.
- Even though, in accordance with the law of conservation of energy, energy can never be created or destroyed, energy that can be used to do work is always partly converted to less useful forms, such as waste heat to the environment, in all of our uses of energy for practical purposes.

Conceptual Questions

Exercise:

Problem:

What is the difference between energy conservation and the law of conservation of energy? Give some examples of each.

Exercise:

Problem:

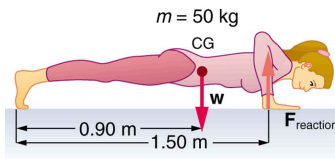
If the efficiency of a coal-fired electrical generating plant is 35%, then what do we mean when we say that energy is a conserved quantity?

Problems & Exercises

Exercise:

Problem: Integrated Concepts

(a) Calculate the force the woman in [\[link\]](#) exerts to do a push-up at constant speed, taking all data to be known to three digits. (b) How much work does she do if her center of mass rises 0.240 m? (c) What is her useful power output if she does 25 push-ups in 1 min? (Should work done lowering her body be included? See the discussion of useful work in [Work, Energy, and Power in Humans](#).



Forces involved in doing push-ups. The woman's weight acts as a force exerted downward on her center of gravity (CG).

Solution:

- (a) 294 N
- (b) 118 J
- (c) 49.0 W

Exercise:

Problem: Integrated Concepts

A 75.0-kg cross-country skier is climbing a 3.0° slope at a constant speed of 2.00 m/s and encounters air resistance of 25.0 N. Find his power output for work done against the gravitational force and air resistance. (b) What average force does he exert backward on the snow to accomplish this? (c) If he continues to exert this force and to experience the same air resistance when he reaches a level area, how long will it take him to reach a velocity of 10.0 m/s?

Exercise:**Problem: Integrated Concepts**

The 70.0-kg swimmer in [\[link\]](#) starts a race with an initial velocity of 1.25 m/s and exerts an average force of 80.0 N backward with his arms during each 1.80 m long stroke. (a) What is his initial acceleration if water resistance is 45.0 N? (b) What is the subsequent average resistance force from the water during the 5.00 s it takes him to reach his top velocity of 2.50 m/s? (c) Discuss whether water resistance seems to increase linearly with velocity.

Solution:

(a) 0.500 m/s^2

(b) 62.5 N

(c) Assuming the acceleration of the swimmer decreases linearly with time over the 5.00 s interval, the frictional force must therefore be increasing linearly with time, since $f = F - ma$. If the acceleration decreases linearly with time, the velocity will contain a term dependent on time squared (t^2). Therefore, the water resistance will not depend linearly on the velocity.

Exercise:**Problem: Integrated Concepts**

A toy gun uses a spring with a force constant of 300 N/m to propel a 10.0-g steel ball. If the spring is compressed 7.00 cm and friction is negligible: (a) How much force is needed to compress the spring? (b) To what maximum height can the ball be shot? (c) At what angles above the horizontal may a child aim to hit a target 3.00 m away at the same height as the gun? (d) What is the gun's maximum range on level ground?

Exercise:**Problem: Integrated Concepts**

(a) What force must be supplied by an elevator cable to produce an acceleration of 0.800 m/s^2 against a 200-N frictional force, if the mass of the loaded elevator is 1500 kg? (b) How much work is done by the cable in lifting the elevator 20.0 m? (c) What is the final speed of the elevator if it starts from rest? (d) How much work went into thermal energy?

Solution:

(a) $16.1 \times 10^3 \text{ N}$

(b) $3.22 \times 10^5 \text{ J}$

(c) 5.66 m/s

(d) 4.00 kJ

Exercise:**Problem: Unreasonable Results**

A car advertisement claims that its 900-kg car accelerated from rest to 30.0 m/s and drove 100 km, gaining 3.00 km in altitude, on 1.0 gal of gasoline. The average force of friction including air resistance was 700 N. Assume all values are known to three significant figures. (a) Calculate the car's efficiency. (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Exercise:**Problem: Unreasonable Results**

Body fat is metabolized, supplying 9.30 kcal/g, when dietary intake is less than needed to fuel metabolism. The manufacturers of an exercise bicycle claim that you can lose 0.500 kg of fat per day by vigorously exercising for 2.00 h per day on their machine. (a) How many kcal are supplied by the metabolization of 0.500 kg of fat? (b) Calculate the kcal/min that you would have to utilize to metabolize fat at the rate of 0.500 kg in 2.00 h. (c) What is unreasonable about the results? (d) Which premise is unreasonable, or which premises are inconsistent?

Solution:

(a) 4.65×10^3 kcal

(b) 38.8 kcal/min

(c) This power output is higher than the highest value on [\[link\]](#), which is about 35 kcal/min (corresponding to 2415 watts) for sprinting.

(d) It would be impossible to maintain this power output for 2 hours (imagine sprinting for 2 hours!).

Exercise:**Problem: Construct Your Own Problem**

Consider a person climbing and descending stairs. Construct a problem in which you calculate the long-term rate at which stairs can be climbed considering the mass of the person, his ability to generate power with his legs, and the height of a single stair step. Also consider why the same person can descend stairs at a faster rate for a nearly unlimited time in spite of the fact that very similar forces are exerted going down as going up. (This points to a fundamentally different process for descending versus climbing stairs.)

Exercise:**Problem: Construct Your Own Problem**

Consider humans generating electricity by pedaling a device similar to a stationary bicycle. Construct a problem in which you determine the number of people it would take to replace a large electrical generation facility. Among the things to consider are the power output that is reasonable using the legs, rest time, and the need for electricity 24 hours per day. Discuss the practical implications of your results.

Exercise:**Problem: Integrated Concepts**

A 105-kg basketball player crouches down 0.400 m while waiting to jump. After exerting a force on the floor through this 0.400 m, his feet leave the floor and his center of gravity rises 0.950 m above its normal standing erect position. (a) Using energy considerations, calculate his velocity when he leaves the floor. (b) What average force did he exert on the floor? (Do not neglect the force to support his weight as well as that to accelerate him.) (c) What was his power output during the acceleration phase?

Solution:

(a) 4.32 m/s

(b) 3.47×10^3 N

(c) 8.93 kW

Glossary

renewable forms of energy

those sources that cannot be used up, such as water, wind, solar, and biomass

fossil fuels

oil, natural gas, and coal

Introduction to Linear Momentum and Collisions

class="introduction"

"Each
rugby
player has
great
momentum
, which will
affect the
outcome of
their
collisions
with each
other and
the ground.
(credit:
vjpaul,
Flickr)"



We use the term momentum in various ways in everyday language, and most of these ways are consistent with its precise scientific definition. We speak of sports teams or politicians gaining and maintaining the momentum to win. We also recognize that momentum has something to do with collisions. For example, looking at the rugby players in the photograph colliding and falling to the ground, we expect their momenta to have great effects in the resulting collisions. Generally, momentum implies a tendency to continue on course—to move in the same direction—and is associated with great mass and speed.

Momentum, like energy, is important because it is conserved. Only a few physical quantities are conserved in nature, and studying them yields fundamental insight into how nature works, as we shall see in our study of momentum.

Linear Momentum and Force

- Define linear momentum.
- Explain the relationship between momentum and force.
- State Newton's second law of motion in terms of momentum.
- Calculate momentum given mass and velocity.

Linear Momentum

The scientific definition of linear momentum is consistent with most people's intuitive understanding of momentum: a large, fast-moving object has greater momentum than a smaller, slower object. **Linear momentum** is defined as the product of a system's mass multiplied by its velocity. In symbols, linear momentum is expressed as

Equation:

$$\mathbf{p} = m\mathbf{v}.$$

Momentum is directly proportional to the object's mass and also its velocity. Thus the greater an object's mass or the greater its velocity, the greater its momentum. Momentum \mathbf{p} is a vector having the same direction as the velocity \mathbf{v} . The SI unit for momentum is $\text{kg} \cdot \text{m/s}$.

Note:

Linear Momentum

Linear momentum is defined as the product of a system's mass multiplied by its velocity:

Equation:

$$\mathbf{p} = m\mathbf{v}.$$

Example:

Calculating Momentum: A Football Player and a Football

(a) Calculate the momentum of a 110-kg football player running at 8.00 m/s. (b) Compare the player's momentum with the momentum of a hard-thrown 0.410-kg football that has a speed of 25.0 m/s.

Strategy

No information is given regarding direction, and so we can calculate only the magnitude of the momentum, p . (As usual, a symbol that is in italics is a magnitude, whereas one that is italicized, boldfaced, and has an arrow is a vector.) In both parts of this example, the magnitude of momentum can be calculated directly from the definition of momentum given in the equation, which becomes

Equation:

$$p = mv$$

when only magnitudes are considered.

Solution for (a)

To determine the momentum of the player, substitute the known values for the player's mass and speed into the equation.

Equation:

$$p_{\text{player}} = (110 \text{ kg})(8.00 \text{ m/s}) = 880 \text{ kg} \cdot \text{m/s}$$

Solution for (b)

To determine the momentum of the ball, substitute the known values for the ball's mass and speed into the equation.

Equation:

$$p_{\text{ball}} = (0.410 \text{ kg})(25.0 \text{ m/s}) = 10.3 \text{ kg} \cdot \text{m/s}$$

The ratio of the player's momentum to that of the ball is

Equation:

$$\frac{p_{\text{player}}}{p_{\text{ball}}} = \frac{880}{10.3} = 85.9.$$

Discussion

Although the ball has greater velocity, the player has a much greater mass. Thus the momentum of the player is much greater than the momentum of the football, as you might guess. As a result, the player's motion is only slightly affected if he catches the ball. We shall quantify what happens in such collisions in terms of momentum in later sections.

Momentum and Newton's Second Law

The importance of momentum, unlike the importance of energy, was recognized early in the development of classical physics. Momentum was deemed so important that it was called the “quantity of motion.” Newton actually stated his **second law of motion** in terms of momentum: The net external force equals the change in momentum of a system divided by the time over which it changes. Using symbols, this law is

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t},$$

where \mathbf{F}_{net} is the net external force, $\Delta \mathbf{p}$ is the change in momentum, and Δt is the change in time.

Note:

Newton's Second Law of Motion in Terms of Momentum

The net external force equals the change in momentum of a system divided by the time over which it changes.

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$$

Note:**Making Connections: Force and Momentum**

Force and momentum are intimately related. Force acting over time can change momentum, and Newton's second law of motion, can be stated in its most broadly applicable form in terms of momentum. Momentum continues to be a key concept in the study of atomic and subatomic particles in quantum mechanics.

This statement of Newton's second law of motion includes the more familiar $\mathbf{F}_{\text{net}} = m\mathbf{a}$ as a special case. We can derive this form as follows. First, note that the change in momentum $\Delta\mathbf{p}$ is given by

Equation:

$$\Delta\mathbf{p} = \Delta(m\mathbf{v}).$$

If the mass of the system is constant, then

Equation:

$$\Delta(m\mathbf{v}) = m\Delta\mathbf{v}.$$

So that for constant mass, Newton's second law of motion becomes

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta\mathbf{p}}{\Delta t} = \frac{m\Delta\mathbf{v}}{\Delta t}.$$

Because $\frac{\Delta\mathbf{v}}{\Delta t} = \mathbf{a}$, we get the familiar equation

Equation:

$$\mathbf{F}_{\text{net}} = m\mathbf{a}$$

when the mass of the system is constant.

Newton's second law of motion stated in terms of momentum is more generally applicable because it can be applied to systems where the mass is changing, such as rockets, as well as to systems of constant mass. We will consider systems with varying mass in some detail; however, the relationship between momentum and force remains useful when mass is constant, such as in the following example.

Example:

Calculating Force: Venus Williams' Racquet

During the 2007 French Open, Venus Williams hit the fastest recorded serve in a premier women's match, reaching a speed of 58 m/s (209 km/h). What is the average force exerted on the 0.057-kg tennis ball by Venus Williams' racquet, assuming that the ball's speed just after impact is 58 m/s, that the initial horizontal component of the velocity before impact is negligible, and that the ball remained in contact with the racquet for 5.0 ms (milliseconds)?

Strategy

This problem involves only one dimension because the ball starts from having no horizontal velocity component before impact. Newton's second law stated in terms of momentum is then written as

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}.$$

As noted above, when mass is constant, the change in momentum is given by

Equation:

$$\Delta p = m\Delta v = m(v_f - v_i).$$

In this example, the velocity just after impact and the change in time are given; thus, once Δp is calculated, $F_{\text{net}} = \frac{\Delta p}{\Delta t}$ can be used to find the force.

Solution

To determine the change in momentum, substitute the values for the initial and final velocities into the equation above.

Equation:

$$\begin{aligned}\Delta p &= m(v_f - v_i) \\ &= (0.057 \text{ kg})(58 \text{ m/s} - 0 \text{ m/s}) \\ &= 3.306 \text{ kg} \cdot \text{m/s} \approx 3.3 \text{ kg} \cdot \text{m/s}\end{aligned}$$

Now the magnitude of the net external force can be determined by using

$$F_{\text{net}} = \frac{\Delta p}{\Delta t}:$$

Equation:

$$\begin{aligned}F_{\text{net}} &= \frac{\Delta p}{\Delta t} = \frac{3.306 \text{ kg} \cdot \text{m/s}}{5.0 \times 10^{-3} \text{ s}} \\ &= 661 \text{ N} \approx 660 \text{ N},\end{aligned}$$

where we have retained only two significant figures in the final step.

Discussion

This quantity was the average force exerted by Venus Williams' racquet on the tennis ball during its brief impact (note that the ball also experienced the 0.56-N force of gravity, but that force was not due to the racquet). This problem could also be solved by first finding the acceleration and then using $F_{\text{net}} = ma$, but one additional step would be required compared with the strategy used in this example.

Section Summary

- Linear momentum (*momentum* for brevity) is defined as the product of a system's mass multiplied by its velocity.
- In symbols, linear momentum **p** is defined to be

Equation:

$$\mathbf{p} = m\mathbf{v},$$

where m is the mass of the system and \mathbf{v} is its velocity.

- The SI unit for momentum is $\text{kg} \cdot \text{m/s}$.

- Newton's second law of motion in terms of momentum states that the net external force equals the change in momentum of a system divided by the time over which it changes.
- In symbols, Newton's second law of motion is defined to be

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t},$$

\mathbf{F}_{net} is the net external force, $\Delta \mathbf{p}$ is the change in momentum, and Δt is the change time.

Conceptual Questions

Exercise:

Problem:

An object that has a small mass and an object that has a large mass have the same momentum. Which object has the largest kinetic energy?

Exercise:

Problem:

An object that has a small mass and an object that has a large mass have the same kinetic energy. Which mass has the largest momentum?

Exercise:

Problem: Professional Application

Football coaches advise players to block, hit, and tackle with their feet on the ground rather than by leaping through the air. Using the concepts of momentum, work, and energy, explain how a football player can be more effective with his feet on the ground.

Exercise:

Problem:

How can a small force impart the same momentum to an object as a large force?

Problems & Exercises**Exercise:****Problem:**

(a) Calculate the momentum of a 2000-kg elephant charging a hunter at a speed of 7.50 m/s. (b) Compare the elephant's momentum with the momentum of a 0.0400-kg tranquilizer dart fired at a speed of 600 m/s. (c) What is the momentum of the 90.0-kg hunter running at 7.40 m/s after missing the elephant?

Solution:

(a) $1.50 \times 10^4 \text{ kg} \cdot \text{m/s}$

(b) 625 to 1

(c) $6.66 \times 10^2 \text{ kg} \cdot \text{m/s}$

Exercise:**Problem:**

(a) What is the mass of a large ship that has a momentum of $1.60 \times 10^9 \text{ kg} \cdot \text{m/s}$, when the ship is moving at a speed of 48.0 km/h? (b) Compare the ship's momentum to the momentum of a 1100-kg artillery shell fired at a speed of 1200 m/s.

Exercise:

Problem:

(a) At what speed would a 2.00×10^4 -kg airplane have to fly to have a momentum of $1.60 \times 10^9 \text{ kg} \cdot \text{m/s}$ (the same as the ship's momentum in the problem above)? (b) What is the plane's momentum when it is taking off at a speed of 60.0 m/s ? (c) If the ship is an aircraft carrier that launches these airplanes with a catapult, discuss the implications of your answer to (b) as it relates to recoil effects of the catapult on the ship.

Solution:

(a) $8.00 \times 10^4 \text{ m/s}$

(b) $1.20 \times 10^6 \text{ kg} \cdot \text{m/s}$

(c) Because the momentum of the airplane is 3 orders of magnitude smaller than of the ship, the ship will not recoil very much. The recoil would be -0.0100 m/s , which is probably not noticeable.

Exercise:**Problem:**

(a) What is the momentum of a garbage truck that is $1.20 \times 10^4 \text{ kg}$ and is moving at 10.0 m/s ? (b) At what speed would an 8.00 -kg trash can have the same momentum as the truck?

Exercise:**Problem:**

A runaway train car that has a mass of $15,000 \text{ kg}$ travels at a speed of 5.4 m/s down a track. Compute the time required for a force of 1500 N to bring the car to rest.

Solution:

54 s

Exercise:**Problem:**

The mass of Earth is 5.972×10^{24} kg and its orbital radius is an average of 1.496×10^{11} m. Calculate its linear momentum.

Glossary

linear momentum

the product of mass and velocity

second law of motion

physical law that states that the net external force equals the change in momentum of a system divided by the time over which it changes

Impulse

- Define impulse.
- Describe effects of impulses in everyday life.
- Determine the average effective force using graphical representation.
- Calculate average force and impulse given mass, velocity, and time.

The effect of a force on an object depends on how long it acts, as well as how great the force is. In [\[link\]](#), a very large force acting for a short time had a great effect on the momentum of the tennis ball. A small force could cause the same **change in momentum**, but it would have to act for a much longer time. For example, if the ball were thrown upward, the gravitational force (which is much smaller than the tennis racquet's force) would eventually reverse the momentum of the ball. Quantitatively, the effect we are talking about is the change in momentum $\Delta \mathbf{p}$.

By rearranging the equation $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$ to be

Equation:

$$\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t,$$

we can see how the change in momentum equals the average net external force multiplied by the time this force acts. The quantity $\mathbf{F}_{\text{net}} \Delta t$ is given the name **impulse**. Impulse is the same as the change in momentum.

Note:

Impulse: Change in Momentum

Change in momentum equals the average net external force multiplied by the time this force acts.

Equation:

$$\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t$$

The quantity $\mathbf{F}_{\text{net}} \Delta t$ is given the name impulse.

There are many ways in which an understanding of impulse can save lives, or at least limbs. The dashboard padding in a car, and certainly the airbags, allow the net force on the occupants in the car to act over a much longer time when there is a sudden stop. The momentum change is the same for an occupant, whether an air bag is deployed or not, but the force (to bring the occupant to a stop) will be much less if it acts over a larger time. Cars today have many plastic components. One advantage of plastics is their lighter weight, which results in better gas mileage. Another advantage is that a car will crumple in a collision, especially in the event of a head-on collision. A longer collision time means the force on the car will be less. Deaths during car races decreased dramatically when the rigid frames of racing cars were replaced with parts that could crumple or collapse in the event of an accident.

Bones in a body will fracture if the force on them is too large. If you jump onto the floor from a table, the force on your legs can be immense if you land stiff-legged on a hard surface. Rolling on the ground after jumping from the table, or landing with a parachute, extends the time over which the force (on you from the ground) acts.

Example:

Calculating Magnitudes of Impulses: Two Billiard Balls Striking a Rigid Wall

Two identical billiard balls strike a rigid wall with the same speed, and are reflected without any change of speed. The first ball strikes perpendicular to the wall. The second ball strikes the wall at an angle of 30° from the perpendicular, and bounces off at an angle of 30° from perpendicular to the wall.

- (a) Determine the direction of the force on the wall due to each ball.
- (b) Calculate the ratio of the magnitudes of impulses on the two balls by the wall.

Strategy for (a)

In order to determine the force on the wall, consider the force on the ball due to the wall using Newton's second law and then apply Newton's third law to determine the direction. Assume the x -axis to be normal to the wall and to be positive in the initial direction of motion. Choose the y -axis to be

along the wall in the plane of the second ball's motion. The momentum direction and the velocity direction are the same.

Solution for (a)

The first ball bounces directly into the wall and exerts a force on it in the $+x$ direction. Therefore the wall exerts a force on the ball in the $-x$ direction. The second ball continues with the same momentum component in the y direction, but reverses its x -component of momentum, as seen by sketching a diagram of the angles involved and keeping in mind the proportionality between velocity and momentum.

These changes mean the change in momentum for both balls is in the $-x$ direction, so the force of the wall on each ball is along the $-x$ direction.

Strategy for (b)

Calculate the change in momentum for each ball, which is equal to the impulse imparted to the ball.

Solution for (b)

Let u be the speed of each ball before and after collision with the wall, and m the mass of each ball. Choose the x -axis and y -axis as previously described, and consider the change in momentum of the first ball which strikes perpendicular to the wall.

Equation:

$$p_{xi} = mu; p_{yi} = 0$$

Equation:

$$p_{xf} = -mu; p_{yf} = 0$$

Impulse is the change in momentum vector. Therefore the x -component of impulse is equal to $-2mu$ and the y -component of impulse is equal to zero.

Now consider the change in momentum of the second ball.

Equation:

$$p_{xi} = mu \cos 30^\circ; p_{yi} = -mu \sin 30^\circ$$

Equation:

$$p_{xf} = -mu \cos 30^\circ; p_{yf} = -mu \sin 30^\circ$$

It should be noted here that while p_x changes sign after the collision, p_y does not. Therefore the x -component of impulse is equal to $-2mu \cos 30^\circ$ and the y -component of impulse is equal to zero.

The ratio of the magnitudes of the impulse imparted to the balls is

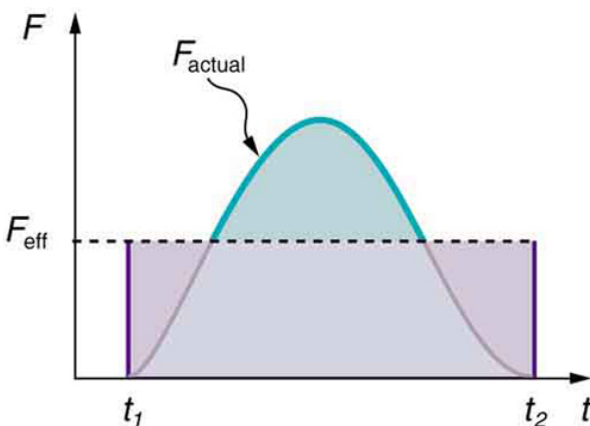
Equation:

$$\frac{2mu}{2mu \cos 30^\circ} = \frac{2}{\sqrt{3}} = 1.155.$$

Discussion

The direction of impulse and force is the same as in the case of (a); it is normal to the wall and along the negative x -direction. Making use of Newton's third law, the force on the wall due to each ball is normal to the wall along the positive x -direction.

Our definition of impulse includes an assumption that the force is constant over the time interval Δt . *Forces are usually not constant.* Forces vary considerably even during the brief time intervals considered. It is, however, possible to find an average effective force F_{eff} that produces the same result as the corresponding time-varying force. [\[link\]](#) shows a graph of what an actual force looks like as a function of time for a ball bouncing off the floor. The area under the curve has units of momentum and is equal to the impulse or change in momentum between times t_1 and t_2 . That area is equal to the area inside the rectangle bounded by F_{eff} , t_1 , and t_2 . Thus the impulses and their effects are the same for both the actual and effective forces.



A graph of force versus time with time along the x -axis and force along the y -axis for an actual force and an equivalent effective force. The areas under the two curves are equal.

Note:

Making Connections: Take-Home Investigation—Hand Movement and Impulse

Try catching a ball while “giving” with the ball, pulling your hands toward your body. Then, try catching a ball while keeping your hands still. Hit water in a tub with your full palm. After the water has settled, hit the water again by diving your hand with your fingers first into the water. (Your full palm represents a swimmer doing a belly flop and your diving hand represents a swimmer doing a dive.) Explain what happens in each case and why. Which orientations would you advise people to avoid and why?

Note:

Making Connections: Constant Force and Constant Acceleration

The assumption of a constant force in the definition of impulse is analogous to the assumption of a constant acceleration in kinematics. In both cases, nature is adequately described without the use of calculus.

Section Summary

- Impulse, or change in momentum, equals the average net external force multiplied by the time this force acts:

Equation:

$$\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t.$$

- Forces are usually not constant over a period of time.

Conceptual Questions

Exercise:

Problem: Professional Application

Explain in terms of impulse how padding reduces forces in a collision. State this in terms of a real example, such as the advantages of a carpeted vs. tile floor for a day care center.

Exercise:

Problem:

While jumping on a trampoline, sometimes you land on your back and other times on your feet. In which case can you reach a greater height and why?

Exercise:

Problem: Professional Application

Tennis racquets have “sweet spots.” If the ball hits a sweet spot then the player's arm is not jarred as much as it would be otherwise. Explain why this is the case.

Problems & Exercises

Exercise:

Problem:

A bullet is accelerated down the barrel of a gun by hot gases produced in the combustion of gun powder. What is the average force exerted on a 0.0300-kg bullet to accelerate it to a speed of 600 m/s in a time of 2.00 ms (milliseconds)?

Solution:

$$9.00 \times 10^3 \text{ N}$$

Exercise:**Problem: Professional Application**

A car moving at 10 m/s crashes into a tree and stops in 0.26 s. Calculate the force the seat belt exerts on a passenger in the car to bring him to a halt. The mass of the passenger is 70 kg.

Exercise:**Problem:**

A person slaps her leg with her hand, bringing her hand to rest in 2.50 milliseconds from an initial speed of 4.00 m/s. (a) What is the average force exerted on the leg, taking the effective mass of the hand and forearm to be 1.50 kg? (b) Would the force be any different if the woman clapped her hands together at the same speed and brought them to rest in the same time? Explain why or why not.

Solution:

a) $2.40 \times 10^3 \text{ N}$ toward the leg

b) The force on each hand would have the same magnitude as that found in part (a) (but in opposite directions by Newton's third law) because the change in momentum and the time interval are the same.

Exercise:

Problem: Professional Application

A professional boxer hits his opponent with a 1000-N horizontal blow that lasts for 0.150 s. (a) Calculate the impulse imparted by this blow. (b) What is the opponent's final velocity, if his mass is 105 kg and he is motionless in midair when struck near his center of mass? (c) Calculate the recoil velocity of the opponent's 10.0-kg head if hit in this manner, assuming the head does not initially transfer significant momentum to the boxer's body. (d) Discuss the implications of your answers for parts (b) and (c).

Exercise:**Problem: Professional Application**

Suppose a child drives a bumper car head on into the side rail, which exerts a force of 4000 N on the car for 0.200 s. (a) What impulse is imparted by this force? (b) Find the final velocity of the bumper car if its initial velocity was 2.80 m/s and the car plus driver have a mass of 200 kg. You may neglect friction between the car and floor.

Solution:

- a) 800 kg · m/s away from the wall
- b) 1.20 m/s away from the wall

Exercise:**Problem: Professional Application**

One hazard of space travel is debris left by previous missions. There are several thousand objects orbiting Earth that are large enough to be detected by radar, but there are far greater numbers of very small objects, such as flakes of paint. Calculate the force exerted by a 0.100-mg chip of paint that strikes a spacecraft window at a relative speed of 4.00×10^3 m/s, given the collision lasts 6.00×10^{-8} s.

Exercise:**Problem: Professional Application**

A 75.0-kg person is riding in a car moving at 20.0 m/s when the car runs into a bridge abutment. (a) Calculate the average force on the person if he is stopped by a padded dashboard that compresses an average of 1.00 cm. (b) Calculate the average force on the person if he is stopped by an air bag that compresses an average of 15.0 cm.

Solution:

(a) 1.50×10^6 N away from the dashboard

(b) 1.00×10^5 N away from the dashboard

Exercise:**Problem: Professional Application**

Military rifles have a mechanism for reducing the recoil forces of the gun on the person firing it. An internal part recoils over a relatively large distance and is stopped by damping mechanisms in the gun. The larger distance reduces the average force needed to stop the internal part. (a) Calculate the recoil velocity of a 1.00-kg plunger that directly interacts with a 0.0200-kg bullet fired at 600 m/s from the gun. (b) If this part is stopped over a distance of 20.0 cm, what average force is exerted upon it by the gun? (c) Compare this to the force exerted on the gun if the bullet is accelerated to its velocity in 10.0 ms (milliseconds).

Exercise:

Problem:

A cruise ship with a mass of 1.00×10^7 kg strikes a pier at a speed of 0.750 m/s. It comes to rest 6.00 m later, damaging the ship, the pier, and the tugboat captain's finances. Calculate the average force exerted on the pier using the concept of impulse. (Hint: First calculate the time it took to bring the ship to rest.)

Solution:

4.69×10^5 N in the boat's original direction of motion

Exercise:**Problem:**

Calculate the final speed of a 110-kg rugby player who is initially running at 8.00 m/s but collides head-on with a padded goalpost and experiences a backward force of 1.76×10^4 N for 5.50×10^{-2} s.

Exercise:**Problem:**

Water from a fire hose is directed horizontally against a wall at a rate of 50.0 kg/s and a speed of 42.0 m/s. Calculate the magnitude of the force exerted on the wall, assuming the water's horizontal momentum is reduced to zero.

Solution:

2.10×10^3 N away from the wall

Exercise:**Problem:**

A 0.450-kg hammer is moving horizontally at 7.00 m/s when it strikes a nail and comes to rest after driving the nail 1.00 cm into a board. (a) Calculate the duration of the impact. (b) What was the average force exerted on the nail?

Exercise:**Problem:**

Starting with the definitions of momentum and kinetic energy, derive an equation for the kinetic energy of a particle expressed as a function of its momentum.

Solution:**Equation:**

$$\begin{aligned}\mathbf{p} &= m\mathbf{v} \Rightarrow p^2 = m^2v^2 \Rightarrow \frac{p^2}{m} = mv^2 \\ \Rightarrow \frac{p^2}{2m} &= \frac{1}{2}mv^2 = \text{KE} \\ KE &= \frac{p^2}{2m}\end{aligned}$$

Exercise:**Problem:**

A ball with an initial velocity of 10 m/s moves at an angle 60° above the $+x$ -direction. The ball hits a vertical wall and bounces off so that it is moving 60° above the $-x$ -direction with the same speed. What is the impulse delivered by the wall?

Exercise:**Problem:**

When serving a tennis ball, a player hits the ball when its velocity is zero (at the highest point of a vertical toss). The racquet exerts a force of 540 N on the ball for 5.00 ms, giving it a final velocity of 45.0 m/s. Using these data, find the mass of the ball.

Solution:

60.0 g

Exercise:

Problem:

A punter drops a ball from rest vertically 1 meter down onto his foot. The ball leaves the foot with a speed of 18 m/s at an angle 55° above the horizontal. What is the impulse delivered by the foot (magnitude and direction)?

Glossary

change in momentum

the difference between the final and initial momentum; the mass times the change in velocity

impulse

the average net external force times the time it acts; equal to the change in momentum

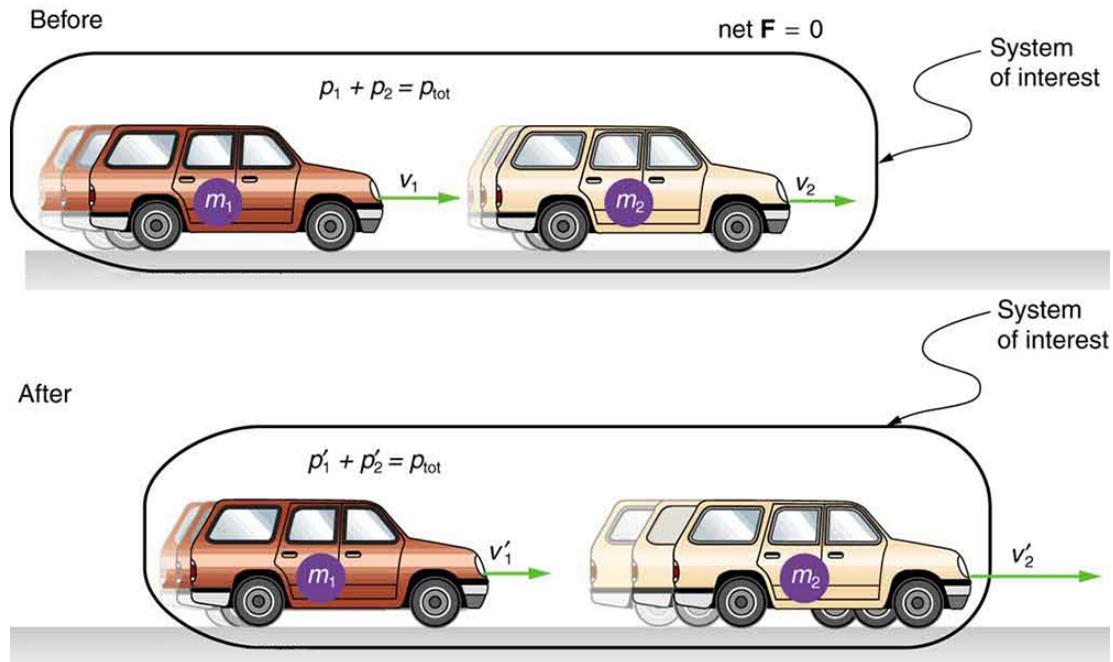
Conservation of Momentum

- Describe the principle of conservation of momentum.
- Derive an expression for the conservation of momentum.
- Explain conservation of momentum with examples.
- Explain the principle of conservation of momentum as it relates to atomic and subatomic particles.

Momentum is an important quantity because it is conserved. Yet it was not conserved in the examples in [Impulse](#) and [Linear Momentum and Force](#), where large changes in momentum were produced by forces acting on the system of interest. Under what circumstances is momentum conserved?

The answer to this question entails considering a sufficiently large system. It is always possible to find a larger system in which total momentum is constant, even if momentum changes for components of the system. If a football player runs into the goalpost in the end zone, there will be a force on him that causes him to bounce backward. However, the Earth also recoils—conserving momentum—because of the force applied to it through the goalpost. Because Earth is many orders of magnitude more massive than the player, its recoil is immeasurably small and can be neglected in any practical sense, but it is real nevertheless.

Consider what happens if the masses of two colliding objects are more similar than the masses of a football player and Earth—for example, one car bumping into another, as shown in [\[link\]](#). Both cars are coasting in the same direction when the lead car (labeled m_2) is bumped by the trailing car (labeled m_1). The only unbalanced force on each car is the force of the collision. (Assume that the effects due to friction are negligible.) Car 1 slows down as a result of the collision, losing some momentum, while car 2 speeds up and gains some momentum. We shall now show that the total momentum of the two-car system remains constant.



A car of mass m_1 moving with a velocity of v_1 bumps into another car of mass m_2 and velocity v_2 that it is following. As a result, the first car slows down to a velocity of v'_1 and the second speeds up to a velocity of v'_2 . The momentum of each car is changed, but the total momentum p_{tot} of the two cars is the same before and after the collision (if you assume friction is negligible).

Using the definition of impulse, the change in momentum of car 1 is given by

Equation:

$$\Delta p_1 = F_1 \Delta t,$$

where F_1 is the force on car 1 due to car 2, and Δt is the time the force acts (the duration of the collision). Intuitively, it seems obvious that the collision time is the same for both cars, but it is only true for objects traveling at ordinary speeds. This assumption must be modified for objects travelling

near the speed of light, without affecting the result that momentum is conserved.

Similarly, the change in momentum of car 2 is

Equation:

$$\Delta p_2 = F_2 \Delta t,$$

where F_2 is the force on car 2 due to car 1, and we assume the duration of the collision Δt is the same for both cars. We know from Newton's third law that $F_2 = -F_1$, and so

Equation:

$$\Delta p_2 = -F_1 \Delta t = -\Delta p_1.$$

Thus, the changes in momentum are equal and opposite, and

Equation:

$$\Delta p_1 + \Delta p_2 = 0.$$

Because the changes in momentum add to zero, the total momentum of the two-car system is constant. That is,

Equation:

$$p_1 + p_2 = \text{constant},$$

Equation:

$$p_1 + p_2 = p'_1 + p'_2,$$

where p'_1 and p'_2 are the momenta of cars 1 and 2 after the collision. (We often use primes to denote the final state.)

This result—that momentum is conserved—has validity far beyond the preceding one-dimensional case. It can be similarly shown that total momentum is conserved for any isolated system, with any number of

objects in it. In equation form, the **conservation of momentum principle** for an isolated system is written

Equation:

$$\mathbf{p}_{\text{tot}} = \text{constant},$$

or

Equation:

$$\mathbf{p}_{\text{tot}} = \mathbf{p}'_{\text{tot}},$$

where \mathbf{p}_{tot} is the total momentum (the sum of the momenta of the individual objects in the system) and \mathbf{p}'_{tot} is the total momentum some time later. (The total momentum can be shown to be the momentum of the center of mass of the system.) An **isolated system** is defined to be one for which the net external force is zero ($\mathbf{F}_{\text{net}} = 0$).

Note:

Conservation of Momentum Principle

Equation:

$$\begin{aligned}\mathbf{p}_{\text{tot}} &= \text{constant} \\ \mathbf{p}_{\text{tot}} &= \mathbf{p}'_{\text{tot}} \text{ (isolated system)}\end{aligned}$$

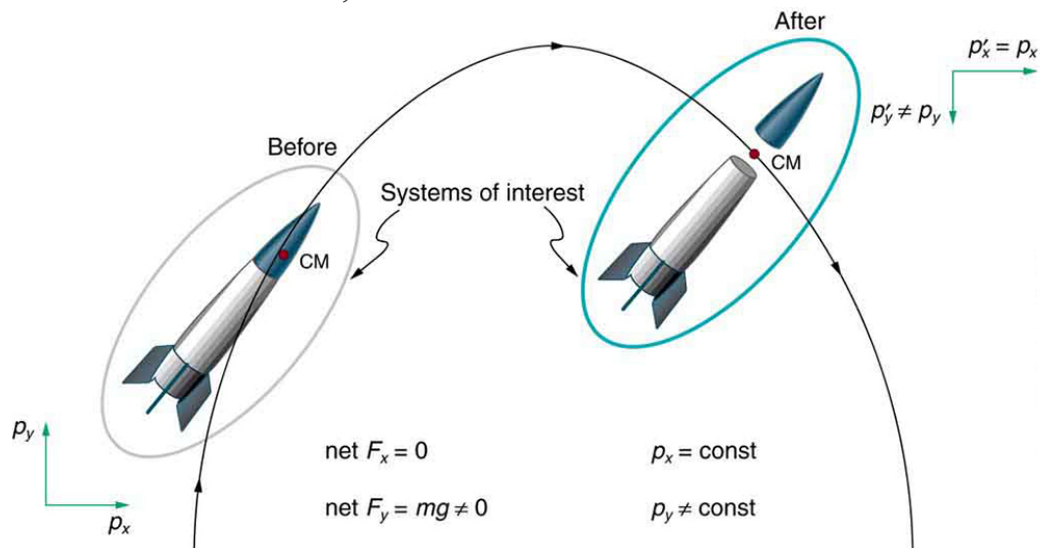
Note:

Isolated System

An isolated system is defined to be one for which the net external force is zero ($\mathbf{F}_{\text{net}} = 0$).

Perhaps an easier way to see that momentum is conserved for an isolated system is to consider Newton's second law in terms of momentum, $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}_{\text{tot}}}{\Delta t}$. For an isolated system, ($\mathbf{F}_{\text{net}} = 0$); thus, $\Delta \mathbf{p}_{\text{tot}} = 0$, and \mathbf{p}_{tot} is constant.

We have noted that the three length dimensions in nature— x , y , and z —are independent, and it is interesting to note that momentum can be conserved in different ways along each dimension. For example, during projectile motion and where air resistance is negligible, momentum is conserved in the horizontal direction because horizontal forces are zero and momentum is unchanged. But along the vertical direction, the net vertical force is not zero and the momentum of the projectile is not conserved. (See [\[link\]](#).) However, if the momentum of the projectile-Earth system is considered in the vertical direction, we find that the total momentum is conserved.



The horizontal component of a projectile's momentum is conserved if air resistance is negligible, even in this case where a space probe separates. The forces causing the separation are internal to the system, so that the net external horizontal force $F_{x-\text{net}}$ is still zero. The vertical component of the momentum is not conserved, because the net vertical force $F_{y-\text{net}}$ is not zero. In the vertical direction, the space probe-Earth system needs to be considered and we find that the total momentum is conserved. The center of mass of the

space probe takes the same path it would if the separation did not occur.

The conservation of momentum principle can be applied to systems as different as a comet striking Earth and a gas containing huge numbers of atoms and molecules. Conservation of momentum is violated only when the net external force is not zero. But another larger system can always be considered in which momentum is conserved by simply including the source of the external force. For example, in the collision of two cars considered above, the two-car system conserves momentum while each one-car system does not.

Note:

Making Connections: Take-Home Investigation—Drop of Tennis Ball and a Basketball

Hold a tennis ball side by side and in contact with a basketball. Drop the balls together. (Be careful!) What happens? Explain your observations. Now hold the tennis ball above and in contact with the basketball. What happened? Explain your observations. What do you think will happen if the basketball ball is held above and in contact with the tennis ball?

Note:

Making Connections: Take-Home Investigation—Two Tennis Balls in a Ballistic Trajectory

Tie two tennis balls together with a string about a foot long. Hold one ball and let the other hang down and throw it in a ballistic trajectory. Explain your observations. Now mark the center of the string with bright ink or attach a brightly colored sticker to it and throw again. What happened? Explain your observations.

Some aquatic animals such as jellyfish move around based on the principles of conservation of momentum. A jellyfish fills its umbrella section with water and then pushes the water out resulting in motion in the opposite direction to that of the jet of water. Squids propel themselves in a

similar manner but, in contrast with jellyfish, are able to control the direction in which they move by aiming their nozzle forward or backward. Typical squids can move at speeds of 8 to 12 km/h.

The ballistocardiograph (BCG) was a diagnostic tool used in the second half of the 20th century to study the strength of the heart. About once a second, your heart beats, forcing blood into the aorta. A force in the opposite direction is exerted on the rest of your body (recall Newton's third law). A ballistocardiograph is a device that can measure this reaction force. This measurement is done by using a sensor (resting on the person) or by using a moving table suspended from the ceiling. This technique can gather information on the strength of the heart beat and the volume of blood passing from the heart. However, the electrocardiogram (ECG or EKG) and the echocardiogram (cardiac ECHO or ECHO; a technique that uses ultrasound to see an image of the heart) are more widely used in the practice of cardiology.

Note:

Making Connections: Conservation of Momentum and Collision

Conservation of momentum is quite useful in describing collisions. Momentum is crucial to our understanding of atomic and subatomic particles because much of what we know about these particles comes from collision experiments.

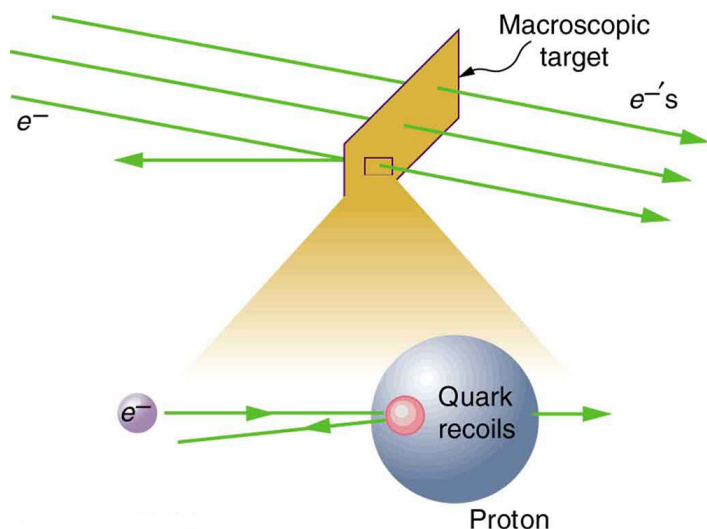
Subatomic Collisions and Momentum

The conservation of momentum principle not only applies to the macroscopic objects, it is also essential to our explorations of atomic and subatomic particles. Giant machines hurl subatomic particles at one another, and researchers evaluate the results by assuming conservation of momentum (among other things).

On the small scale, we find that particles and their properties are invisible to the naked eye but can be measured with our instruments, and models of these subatomic particles can be constructed to describe the results.

Momentum is found to be a property of all subatomic particles including massless particles such as photons that compose light. Momentum being a property of particles hints that momentum may have an identity beyond the description of an object's mass multiplied by the object's velocity. Indeed, momentum relates to wave properties and plays a fundamental role in what measurements are taken and how we take these measurements.

Furthermore, we find that the conservation of momentum principle is valid when considering systems of particles. We use this principle to analyze the masses and other properties of previously undetected particles, such as the nucleus of an atom and the existence of quarks that make up particles of nuclei. [\[link\]](#) below illustrates how a particle scattering backward from another implies that its target is massive and dense. Experiments seeking evidence that **quarks** make up protons (one type of particle that makes up nuclei) scattered high-energy electrons off of protons (nuclei of hydrogen atoms). Electrons occasionally scattered straight backward in a manner that implied a very small and very dense particle makes up the proton—this observation is considered nearly direct evidence of quarks. The analysis was based partly on the same conservation of momentum principle that works so well on the large scale.



A subatomic particle scatters straight backward from a target particle. In experiments seeking evidence for

quarks, electrons were observed to occasionally scatter straight backward from a proton.

Section Summary

- The conservation of momentum principle is written
Equation:

$$\mathbf{p}_{\text{tot}} = \text{constant}$$

or

Equation:

$$\mathbf{p}_{\text{tot}} = \mathbf{p}'_{\text{tot}} \text{ (isolated system),}$$

\mathbf{p}_{tot} is the initial total momentum and \mathbf{p}'_{tot} is the total momentum some time later.

- An isolated system is defined to be one for which the net external force is zero ($\mathbf{F}_{\text{net}} = 0$).
- During projectile motion and where air resistance is negligible, momentum is conserved in the horizontal direction because horizontal forces are zero.
- Conservation of momentum applies only when the net external force is zero.
- The conservation of momentum principle is valid when considering systems of particles.

Conceptual Questions

Exercise:

Problem: Professional Application

If you dive into water, you reach greater depths than if you do a belly flop. Explain this difference in depth using the concept of conservation of energy. Explain this difference in depth using what you have learned in this chapter.

Exercise:

Problem: Under what circumstances is momentum conserved?

Exercise:

Problem:

Can momentum be conserved for a system if there are external forces acting on the system? If so, under what conditions? If not, why not?

Exercise:

Problem:

Momentum for a system can be conserved in one direction while not being conserved in another. What is the angle between the directions? Give an example.

Exercise:

Problem: Professional Application

Explain in terms of momentum and Newton's laws how a car's air resistance is due in part to the fact that it pushes air in its direction of motion.

Exercise:

Problem:

Can objects in a system have momentum while the momentum of the system is zero? Explain your answer.

Exercise:

Problem:

Must the total energy of a system be conserved whenever its momentum is conserved? Explain why or why not.

Problems & Exercises**Exercise:****Problem: Professional Application**

Train cars are coupled together by being bumped into one another. Suppose two loaded train cars are moving toward one another, the first having a mass of 150,000 kg and a velocity of 0.300 m/s, and the second having a mass of 110,000 kg and a velocity of -0.120 m/s. (The minus indicates direction of motion.) What is their final velocity?

Solution:

0.122 m/s

Exercise:**Problem:**

Suppose a clay model of a koala bear has a mass of 0.200 kg and slides on ice at a speed of 0.750 m/s. It runs into another clay model, which is initially motionless and has a mass of 0.350 kg. Both being soft clay, they naturally stick together. What is their final velocity?

Exercise:**Problem: Professional Application**

Consider the following question: *A car moving at 10 m/s crashes into a tree and stops in 0.26 s. Calculate the force the seatbelt exerts on a passenger in the car to bring him to a halt. The mass of the passenger is 70 kg.* Would the answer to this question be different if the car with

the 70-kg passenger had collided with a car that has a mass equal to and is traveling in the opposite direction and at the same speed? Explain your answer.

Solution:

In a collision with an identical car, momentum is conserved. Afterwards $v_f = 0$ for both cars. The change in momentum will be the same as in the crash with the tree. However, the force on the body is not determined since the time is not known. A padded stop will reduce injurious force on body.

Exercise:

Problem:

What is the velocity of a 900-kg car initially moving at 30.0 m/s, just after it hits a 150-kg deer initially running at 12.0 m/s in the same direction? Assume the deer remains on the car.

Exercise:

Problem:

A 1.80-kg falcon catches a 0.650-kg dove from behind in midair. What is their velocity after impact if the falcon's velocity is initially 28.0 m/s and the dove's velocity is 7.00 m/s in the same direction?

Solution:

22.4 m/s in the same direction as the original motion

Glossary

conservation of momentum principle

when the net external force is zero, the total momentum of the system is conserved or constant

isolated system

a system in which the net external force is zero

quark

fundamental constituent of matter and an elementary particle

Elastic Collisions in One Dimension

- Describe an elastic collision of two objects in one dimension.
- Define internal kinetic energy.
- Derive an expression for conservation of internal kinetic energy in a one dimensional collision.
- Determine the final velocities in an elastic collision given masses and initial velocities.

Let us consider various types of two-object collisions. These collisions are the easiest to analyze, and they illustrate many of the physical principles involved in collisions. The conservation of momentum principle is very useful here, and it can be used whenever the net external force on a system is zero.

We start with the elastic collision of two objects moving along the same line—a one-dimensional problem. An **elastic collision** is one that also conserves internal kinetic energy. **Internal kinetic energy** is the sum of the kinetic energies of the objects in the system. [\[link\]](#) illustrates an elastic collision in which internal kinetic energy and momentum are conserved.

Truly elastic collisions can only be achieved with subatomic particles, such as electrons striking nuclei. Macroscopic collisions can be very nearly, but not quite, elastic—some kinetic energy is always converted into other forms of energy such as heat transfer due to friction and sound. One macroscopic collision that is nearly elastic is that of two steel blocks on ice. Another nearly elastic collision is that between two carts with spring bumpers on an air track. Icy surfaces and air tracks are nearly frictionless, more readily allowing nearly elastic collisions on them.

Note:

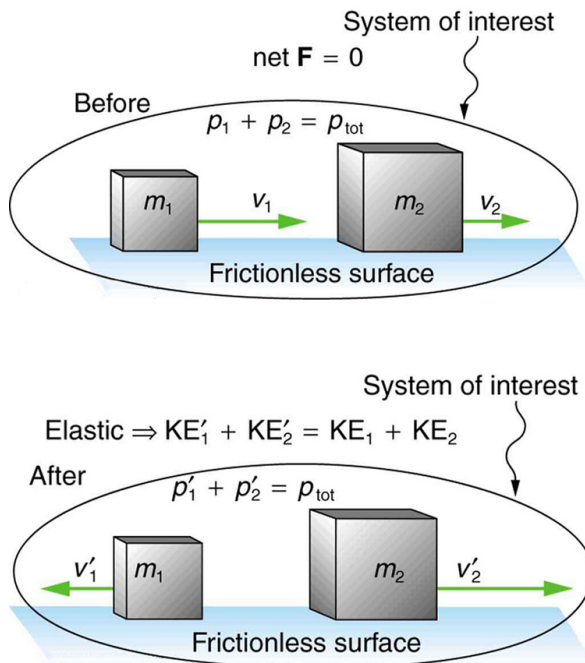
Elastic Collision

An **elastic collision** is one that conserves internal kinetic energy.

Note:

Internal Kinetic Energy

Internal kinetic energy is the sum of the kinetic energies of the objects in the system.



An elastic one-dimensional
two-object collision.
Momentum and internal kinetic
energy are conserved.

Now, to solve problems involving one-dimensional elastic collisions between two objects we can use the equations for conservation of momentum and conservation of internal kinetic energy. First, the equation for conservation of momentum for two objects in a one-dimensional collision is

Equation:

$$p_1 + p_2 = p'_1 + p'_2 \quad (F_{\text{net}} = 0)$$

or

Equation:

$$m_1v_1 + m_2v_2 = m_1v'_1 + m_2v'_2 \quad (F_{\text{net}} = 0),$$

where the primes (') indicate values after the collision. By definition, an elastic collision conserves internal kinetic energy, and so the sum of kinetic energies before the collision equals the sum after the collision. Thus,

Equation:

$$\frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 = \frac{1}{2}m_1v'^2_1 + \frac{1}{2}m_2v'^2_2 \quad (\text{two-object elastic collision})$$

expresses the equation for conservation of internal kinetic energy in a one-dimensional collision.

Example:**Calculating Velocities Following an Elastic Collision**

Calculate the velocities of two objects following an elastic collision, given that

Equation:

$$m_1 = 0.500 \text{ kg}, \quad m_2 = 3.50 \text{ kg}, \quad v_1 = 4.00 \text{ m/s}, \text{ and } v_2 = 0.$$

Strategy and Concept

First, visualize what the initial conditions mean—a small object strikes a larger object that is initially at rest. This situation is slightly simpler than the situation shown in [\[link\]](#) where both objects are initially moving. We are asked to find two unknowns (the final velocities v'_1 and v'_2). To find two unknowns, we must use two independent equations. Because this collision is elastic, we can use the above two equations. Both can be simplified by the fact that object 2 is initially at rest, and thus $v_2 = 0$. Once we simplify these equations, we combine them algebraically to solve for the unknowns.

Solution

For this problem, note that $v_2 = 0$ and use conservation of momentum. Thus,

Equation:

$$p_1 = p'_1 + p'_2$$

or

Equation:

$$m_1 v_1 = m_1 v'_1 + m_2 v'_2.$$

Using conservation of internal kinetic energy and that $v_2 = 0$,

Equation:

$$\frac{1}{2} m_1 v_1^2 = \frac{1}{2} m_1 v'^2_1 + \frac{1}{2} m_2 v'^2_2.$$

Solving the first equation (momentum equation) for v'_2 , we obtain

Equation:

$$v'_2 = \frac{m_1}{m_2} (v_1 - v'_1).$$

Substituting this expression into the second equation (internal kinetic energy equation) eliminates the variable v'_2 , leaving only v'_1 as an unknown (the algebra is left as an exercise for the reader). There are two solutions to any quadratic equation; in this example, they are

Equation:

$$v'_1 = 4.00 \text{ m/s}$$

and

Equation:

$$v'_1 = -3.00 \text{ m/s}.$$

As noted when quadratic equations were encountered in earlier chapters, both solutions may or may not be meaningful. In this case, the first solution is the same as the initial condition. The first solution thus represents the situation before the collision and is discarded. The second solution ($v'_1 = -3.00 \text{ m/s}$) is negative, meaning that the first object bounces backward. When this negative value of v'_1 is used to find the velocity of the second object after the collision, we get

Equation:

$$v'_2 = \frac{m_1}{m_2}(v_1 - v'_1) = \frac{0.500 \text{ kg}}{3.50 \text{ kg}}[4.00 - (-3.00)] \text{ m/s}$$

or

Equation:

$$v'_2 = 1.00 \text{ m/s.}$$

Discussion

The result of this example is intuitively reasonable. A small object strikes a larger one at rest and bounces backward. The larger one is knocked forward, but with a low speed. (This is like a compact car bouncing backward off a full-size SUV that is initially at rest.) As a check, try calculating the internal kinetic energy before and after the collision. You will see that the internal kinetic energy is unchanged at 4.00 J. Also check the total momentum before and after the collision; you will find it, too, is unchanged.

The equations for conservation of momentum and internal kinetic energy as written above can be used to describe any one-dimensional elastic collision of two objects. These equations can be extended to more objects if needed.

Note:

Making Connections: Take-Home Investigation—Ice Cubes and Elastic Collision

Find a few ice cubes which are about the same size and a smooth kitchen tabletop or a table with a glass top. Place the ice cubes on the surface several centimeters away from each other. Flick one ice cube toward a stationary ice cube and observe the path and velocities of the ice cubes after the collision. Try to avoid edge-on collisions and collisions with rotating ice cubes. Have you created approximately elastic collisions? Explain the speeds and directions of the ice cubes using momentum.

Note:

PhET Explorations: Collision Lab

Investigate collisions on an air hockey table. Set up your own experiments: vary the number of discs, masses and initial conditions. Is momentum

conserved? Is kinetic energy conserved? Vary the elasticity and see what happens.

https://phet.colorado.edu/sims/collision-lab/collision-lab_en.html

Section Summary

- An elastic collision is one that conserves internal kinetic energy.
- Conservation of kinetic energy and momentum together allow the final velocities to be calculated in terms of initial velocities and masses in one dimensional two-body collisions.

Conceptual Questions

Exercise:

Problem: What is an elastic collision?

Problems & Exercises

Exercise:

Problem:

Two identical objects (such as billiard balls) have a one-dimensional collision in which one is initially motionless. After the collision, the moving object is stationary and the other moves with the same speed as the other originally had. Show that both momentum and kinetic energy are conserved.

Exercise:

Problem: Professional Application

Two manned satellites approach one another at a relative speed of 0.250 m/s, intending to dock. The first has a mass of 4.00×10^3 kg, and the

second a mass of 7.50×10^3 kg. If the two satellites collide elastically rather than dock, what is their final relative velocity?

Solution:

0.250 m/s

Exercise:

Problem:

A 70.0-kg ice hockey goalie, originally at rest, catches a 0.150-kg hockey puck slapped at him at a velocity of 35.0 m/s. Suppose the goalie and the ice puck have an elastic collision and the puck is reflected back in the direction from which it came. What would their final velocities be in this case?

Glossary

elastic collision

a collision that also conserves internal kinetic energy

internal kinetic energy

the sum of the kinetic energies of the objects in a system

Inelastic Collisions in One Dimension

- Define inelastic collision.
- Explain perfectly inelastic collision.
- Apply an understanding of collisions to sports.
- Determine recoil velocity and loss in kinetic energy given mass and initial velocity.

We have seen that in an elastic collision, internal kinetic energy is conserved. An **inelastic collision** is one in which the internal kinetic energy changes (it is not conserved). This lack of conservation means that the forces between colliding objects may remove or add internal kinetic energy. Work done by internal forces may change the forms of energy within a system. For inelastic collisions, such as when colliding objects stick together, this internal work may transform some internal kinetic energy into heat transfer. Or it may convert stored energy into internal kinetic energy, such as when exploding bolts separate a satellite from its launch vehicle.

Note:

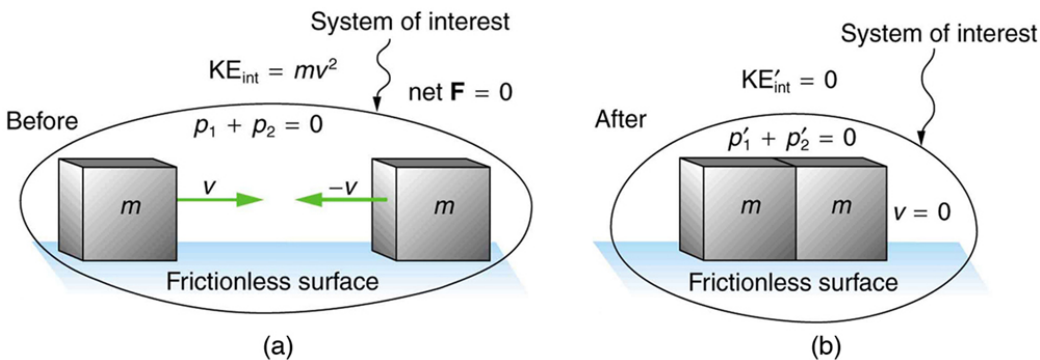
Inelastic Collision

An inelastic collision is one in which the internal kinetic energy changes (it is not conserved).

[\[link\]](#) shows an example of an inelastic collision. Two objects that have equal masses head toward one another at equal speeds and then stick together. Their total internal kinetic energy is initially $\frac{1}{2}mv^2 + \frac{1}{2}mv^2 = mv^2$. The two objects come to rest after sticking together, conserving momentum. But the internal kinetic energy is zero after the collision. A collision in which the objects stick together is sometimes called a **perfectly inelastic collision** because it reduces internal kinetic energy more than does any other type of inelastic collision. In fact, such a collision reduces internal kinetic energy to the minimum it can have while still conserving momentum.

Note:**Perfectly Inelastic Collision**

A collision in which the objects stick together is sometimes called “perfectly inelastic.”



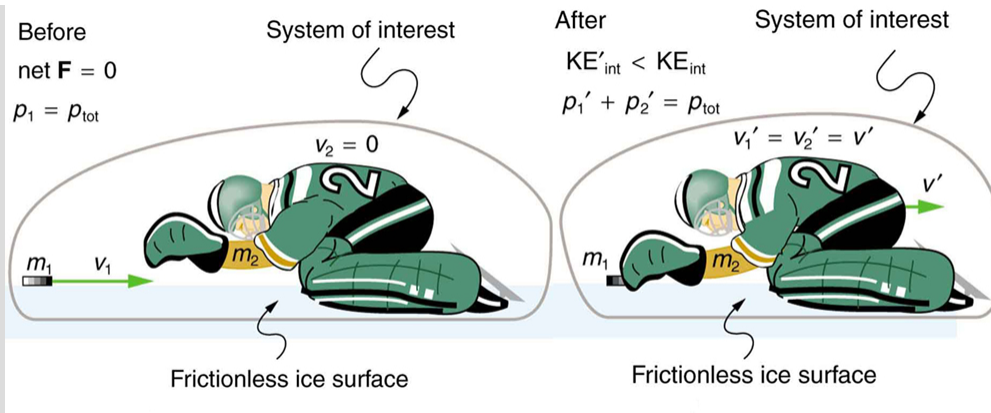
An inelastic one-dimensional two-object collision. Momentum is conserved, but internal kinetic energy is not conserved. (a) Two objects of equal mass initially head directly toward one another at the same speed. (b) The objects stick together (a perfectly inelastic collision), and so their final velocity is zero. The internal kinetic energy of the system changes in any inelastic collision and is reduced to zero in this example.

Example:**Calculating Velocity and Change in Kinetic Energy: Inelastic Collision of a Puck and a Goalie**

(a) Find the recoil velocity of a 70.0-kg ice hockey goalie, originally at rest, who catches a 0.150-kg hockey puck slapped at him at a velocity of 35.0 m/s. (b) How much kinetic energy is lost during the collision?

Assume friction between the ice and the puck-goalie system is negligible.

(See [link](#))



An ice hockey goalie catches a hockey puck and recoils backward. The initial kinetic energy of the puck is almost entirely converted to thermal energy and sound in this inelastic collision.

Strategy

Momentum is conserved because the net external force on the puck-goalie system is zero. We can thus use conservation of momentum to find the final velocity of the puck and goalie system. Note that the initial velocity of the goalie is zero and that the final velocity of the puck and goalie are the same. Once the final velocity is found, the kinetic energies can be calculated before and after the collision and compared as requested.

Solution for (a)

Momentum is conserved because the net external force on the puck-goalie system is zero.

Conservation of momentum is

Equation:

$$p_1 + p_2 = p'_1 + p'_2$$

or

Equation:

$$m_1 v_1 + m_2 v_2 = m_1 v'_1 + m_2 v'_2.$$

Because the goalie is initially at rest, we know $v_2 = 0$. Because the goalie catches the puck, the final velocities are equal, or $v'_1 = v'_2 = v'$. Thus, the

conservation of momentum equation simplifies to

Equation:

$$m_1 v_1 = (m_1 + m_2) v'.$$

Solving for v' yields

Equation:

$$v' = \frac{m_1}{m_1 + m_2} v_1.$$

Entering known values in this equation, we get

Equation:

$$v' = \left(\frac{0.150 \text{ kg}}{0.150 \text{ kg} + 70.0 \text{ kg}} \right) (35.0 \text{ m/s}) = 7.48 \times 10^{-2} \text{ m/s}.$$

Discussion for (a)

This recoil velocity is small and in the same direction as the puck's original velocity, as we might expect.

Solution for (b)

Before the collision, the internal kinetic energy KE_{int} of the system is that of the hockey puck, because the goalie is initially at rest. Therefore, KE_{int} is initially

Equation:

$$\begin{aligned} \text{KE}_{\text{int}} &= \frac{1}{2} m v^2 = \frac{1}{2} (0.150 \text{ kg}) (35.0 \text{ m/s})^2 \\ &= 91.9 \text{ J}. \end{aligned}$$

After the collision, the internal kinetic energy is

Equation:

$$\begin{aligned} \text{KE}'_{\text{int}} &= \frac{1}{2} (m + M) v'^2 = \frac{1}{2} (70.15 \text{ kg}) (7.48 \times 10^{-2} \text{ m/s})^2 \\ &= 0.196 \text{ J}. \end{aligned}$$

The change in internal kinetic energy is thus

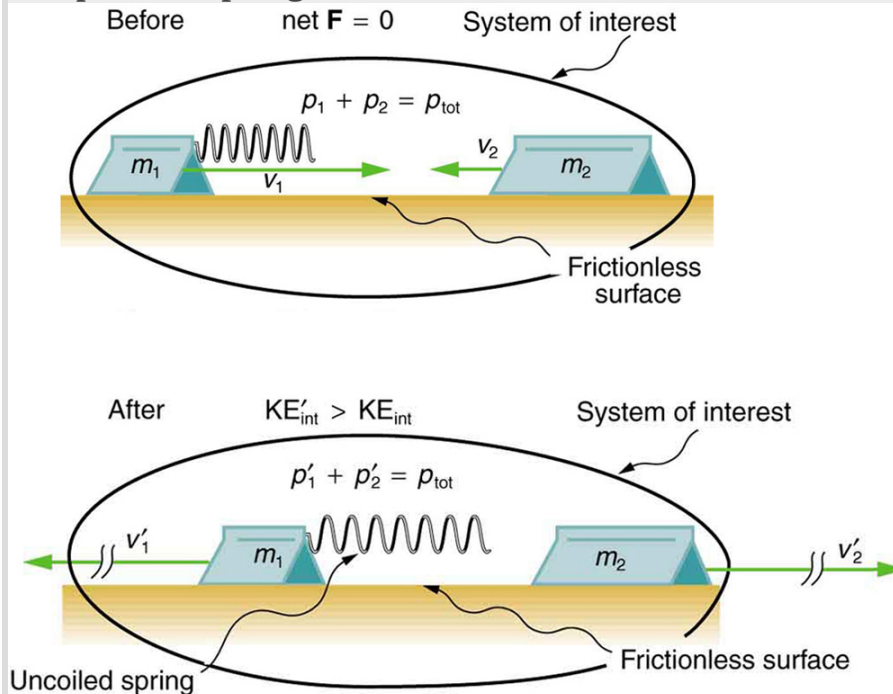
Equation:

$$\begin{aligned} KE_{\text{int}} - KE_{\text{int}} &= 0.196 \text{ J} - 91.9 \text{ J} \\ &= -91.7 \text{ J} \end{aligned}$$

where the minus sign indicates that the energy was lost.

Discussion for (b)

Nearly all of the initial internal kinetic energy is lost in this perfectly inelastic collision. KE_{int} is mostly converted to thermal energy and sound. During some collisions, the objects do not stick together and less of the internal kinetic energy is removed—such as happens in most automobile accidents. Alternatively, stored energy may be converted into internal kinetic energy during a collision. [\[link\]](#) shows a one-dimensional example in which two carts on an air track collide, releasing potential energy from a compressed spring. [\[link\]](#) deals with data from such a collision.



An air track is nearly frictionless, so that momentum is conserved. Motion is one-dimensional. In this collision, examined in [\[link\]](#), the potential energy of a compressed spring is released during the collision and is converted to internal kinetic energy.

Collisions are particularly important in sports and the sporting and leisure industry utilizes elastic and inelastic collisions. Let us look briefly at tennis. Recall that in a collision, it is momentum and not force that is important. So, a heavier tennis racquet will have the advantage over a lighter one. This conclusion also holds true for other sports—a lightweight bat (such as a softball bat) cannot hit a hardball very far.

The location of the impact of the tennis ball on the racquet is also important, as is the part of the stroke during which the impact occurs. A smooth motion results in the maximizing of the velocity of the ball after impact and reduces sports injuries such as tennis elbow. A tennis player tries to hit the ball on the “sweet spot” on the racquet, where the vibration and impact are minimized and the ball is able to be given more velocity. Sports science and technologies also use physics concepts such as momentum and rotational motion and vibrations.

Note:

Take-Home Experiment—Bouncing of Tennis Ball

1. Find a racquet (a tennis, badminton, or other racquet will do). Place the racquet on the floor and stand on the handle. Drop a tennis ball on the strings from a measured height. Measure how high the ball bounces. Now ask a friend to hold the racquet firmly by the handle and drop a tennis ball from the same measured height above the racquet. Measure how high the ball bounces and observe what happens to your friend’s hand during the collision. Explain your observations and measurements.
2. The coefficient of restitution (c) is a measure of the elasticity of a collision between a ball and an object, and is defined as the ratio of the speeds after and before the collision. A perfectly elastic collision has a c of 1. For a ball bouncing off the floor (or a racquet on the floor), c can be shown to be $c = (h/H)^{1/2}$ where h is the height to which the ball bounces and H is the height from which the ball is dropped. Determine c for the cases in Part 1 and for the case of a tennis ball bouncing off a concrete or wooden floor ($c = 0.85$ for new tennis balls used on a tennis court).

Example:**Calculating Final Velocity and Energy Release: Two Carts Collide**

In the collision pictured in [\[link\]](#), two carts collide inelastically. Cart 1 (denoted m_1) carries a spring which is initially compressed. During the collision, the spring releases its potential energy and converts it to internal kinetic energy. The mass of cart 1 and the spring is 0.350 kg, and the cart and the spring together have an initial velocity of 2.00 m/s. Cart 2 (denoted m_2 in [\[link\]](#)) has a mass of 0.500 kg and an initial velocity of -0.500 m/s. After the collision, cart 1 is observed to recoil with a velocity of -4.00 m/s. (a) What is the final velocity of cart 2? (b) How much energy was released by the spring (assuming all of it was converted into internal kinetic energy)?

Strategy

We can use conservation of momentum to find the final velocity of cart 2, because $F_{\text{net}} = 0$ (the track is frictionless and the force of the spring is internal). Once this velocity is determined, we can compare the internal kinetic energy before and after the collision to see how much energy was released by the spring.

Solution for (a)

As before, the equation for conservation of momentum in a two-object system is

Equation:

$$m_1v_1 + m_2v_2 = m_1v'_1 + m_2v'_2.$$

The only unknown in this equation is v'_2 . Solving for v'_2 and substituting known values into the previous equation yields

Equation:

$$\begin{aligned} v'_2 &= \frac{m_1v_1 + m_2v_2 - m_1v'_1}{m_2} \\ &= \frac{(0.350 \text{ kg})(2.00 \text{ m/s}) + (0.500 \text{ kg})(-0.500 \text{ m/s})}{0.500 \text{ kg}} - \frac{(0.350 \text{ kg})(-4.00 \text{ m/s})}{0.500 \text{ kg}} \\ &= 3.70 \text{ m/s.} \end{aligned}$$

Solution for (b)

The internal kinetic energy before the collision is

Equation:

$$\begin{aligned}\text{KE}_{\text{int}} &= \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 \\ &= \frac{1}{2}(0.350 \text{ kg})(2.00 \text{ m/s})^2 + \frac{1}{2}(0.500 \text{ kg})(-0.500 \text{ m/s})^2 \\ &= 0.763 \text{ J.}\end{aligned}$$

After the collision, the internal kinetic energy is

Equation:

$$\begin{aligned}\text{KE}'_{\text{int}} &= \frac{1}{2}m_1v_1'^2 + \frac{1}{2}m_2v_2'^2 \\ &= \frac{1}{2}(0.350 \text{ kg})(-4.00 \text{ m/s})^2 + \frac{1}{2}(0.500 \text{ kg})(3.70 \text{ m/s})^2 \\ &= 6.22 \text{ J.}\end{aligned}$$

The change in internal kinetic energy is thus

Equation:

$$\begin{aligned}\text{KE}'_{\text{int}} - \text{KE}_{\text{int}} &= 6.22 \text{ J} - 0.763 \text{ J} \\ &= 5.46 \text{ J.}\end{aligned}$$

Discussion

The final velocity of cart 2 is large and positive, meaning that it is moving to the right after the collision. The internal kinetic energy in this collision increases by 5.46 J. That energy was released by the spring.

Section Summary

- An inelastic collision is one in which the internal kinetic energy changes (it is not conserved).
- A collision in which the objects stick together is sometimes called perfectly inelastic because it reduces internal kinetic energy more than does any other type of inelastic collision.

- Sports science and technologies also use physics concepts such as momentum and rotational motion and vibrations.

Conceptual Questions

Exercise:

Problem:

What is an inelastic collision? What is a perfectly inelastic collision?

Exercise:

Problem:

Mixed-pair ice skaters performing in a show are standing motionless at arms length just before starting a routine. They reach out, clasp hands, and pull themselves together by only using their arms. Assuming there is no friction between the blades of their skates and the ice, what is their velocity after their bodies meet?

Exercise:

Problem:

A small pickup truck that has a camper shell slowly coasts toward a red light with negligible friction. Two dogs in the back of the truck are moving and making various inelastic collisions with each other and the walls. What is the effect of the dogs on the motion of the center of mass of the system (truck plus entire load)? What is their effect on the motion of the truck?

Problems & Exercises

Exercise:

Problem:

A 0.240-kg billiard ball that is moving at 3.00 m/s strikes the bumper of a pool table and bounces straight back at 2.40 m/s (80% of its original speed). The collision lasts 0.0150 s. (a) Calculate the average force exerted on the ball by the bumper. (b) How much kinetic energy in joules is lost during the collision? (c) What percent of the original energy is left?

Solution:

(a) 86.4 N perpendicularly away from the bumper

(b) 0.389 J

(c) 64.0%

Exercise:**Problem:**

During an ice show, a 60.0-kg skater leaps into the air and is caught by an initially stationary 75.0-kg skater. (a) What is their final velocity assuming negligible friction and that the 60.0-kg skater's original horizontal velocity is 4.00 m/s? (b) How much kinetic energy is lost?

Exercise:**Problem: Professional Application**

Using mass and speed data from [\[link\]](#) and assuming that the football player catches the ball with his feet off the ground with both of them moving horizontally, calculate: (a) the final velocity if the ball and player are going in the same direction and (b) the loss of kinetic energy in this case. (c) Repeat parts (a) and (b) for the situation in which the ball and the player are going in opposite directions. Might the loss of kinetic energy be related to how much it hurts to catch the pass?

Solution:

(a) 8.06 m/s

(b) -56.0 J

(c)(i) 7.88 m/s; (ii) -223 J

Exercise:

Problem:

A battleship that is 6.00×10^7 kg and is originally at rest fires a 1100-kg artillery shell horizontally with a velocity of 575 m/s. (a) If the shell is fired straight aft (toward the rear of the ship), there will be negligible friction opposing the ship's recoil. Calculate its recoil velocity. (b) Calculate the increase in internal kinetic energy (that is, for the ship and the shell). This energy is less than the energy released by the gun powder—significant heat transfer occurs.

Exercise:

Problem: Professional Application

Two manned satellites approaching one another, at a relative speed of 0.250 m/s, intending to dock. The first has a mass of 4.00×10^3 kg, and the second a mass of 7.50×10^3 kg. (a) Calculate the final velocity (after docking) by using the frame of reference in which the first satellite was originally at rest. (b) What is the loss of kinetic energy in this inelastic collision? (c) Repeat both parts by using the frame of reference in which the second satellite was originally at rest. Explain why the change in velocity is different in the two frames, whereas the change in kinetic energy is the same in both.

Solution:

(a) 0.163 m/s in the direction of motion of the more massive satellite

(b) 81.6 J

(c) 8.70×10^{-2} m/s in the direction of motion of the less massive satellite, 81.5 J. Because there are no external forces, the velocity of the center of mass of the two-satellite system is unchanged by the collision. The two velocities calculated above are the velocity of the center of mass in each of the two different individual reference frames. The loss in KE is the same in both reference frames because the KE lost to internal forces (heat, friction, etc.) is the same regardless of the coordinate system chosen.

Exercise:

Problem: Professional Application

A 30,000-kg freight car is coasting at 0.850 m/s with negligible friction under a hopper that dumps 110,000 kg of scrap metal into it. (a) What is the final velocity of the loaded freight car? (b) How much kinetic energy is lost?

Exercise:

Problem: Professional Application

Space probes may be separated from their launchers by exploding bolts. (They bolt away from one another.) Suppose a 4800-kg satellite uses this method to separate from the 1500-kg remains of its launcher, and that 5000 J of kinetic energy is supplied to the two parts. What are their subsequent velocities using the frame of reference in which they were at rest before separation?

Solution:

0.704 m/s

−2.25 m/s

Exercise:

Problem:

A 0.0250-kg bullet is accelerated from rest to a speed of 550 m/s in a 3.00-kg rifle. The pain of the rifle's kick is much worse if you hold the gun loosely a few centimeters from your shoulder rather than holding it tightly against your shoulder. (a) Calculate the recoil velocity of the rifle if it is held loosely away from the shoulder. (b) How much kinetic energy does the rifle gain? (c) What is the recoil velocity if the rifle is held tightly against the shoulder, making the effective mass 28.0 kg? (d) How much kinetic energy is transferred to the rifle-shoulder combination? The pain is related to the amount of kinetic energy, which is significantly less in this latter situation. (e) Calculate the momentum of a 110-kg football player running at 8.00 m/s. Compare the player's momentum with the momentum of a hard-thrown 0.410-kg football that has a speed of 25.0 m/s. Discuss its relationship to this problem.

Solution:

(a) 4.58 m/s away from the bullet

(b) 31.5 J

(c) -0.491 m/s

(d) 3.38 J

Exercise:**Problem: Professional Application**

One of the waste products of a nuclear reactor is plutonium-239 (^{239}Pu). This nucleus is radioactive and decays by splitting into a helium-4 nucleus and a uranium-235 nucleus ($^4\text{He} + ^{235}\text{U}$), the latter of which is also radioactive and will itself decay some time later. The energy emitted in the plutonium decay is 8.40×10^{-13} J and is entirely converted to kinetic energy of the helium and uranium nuclei.

The mass of the helium nucleus is 6.68×10^{-27} kg, while that of the uranium is 3.92×10^{-25} kg (note that the ratio of the masses is 4 to 235). (a) Calculate the velocities of the two nuclei, assuming the plutonium nucleus is originally at rest. (b) How much kinetic energy does each nucleus carry away? Note that the data given here are accurate to three digits only.

Exercise:

Problem: Professional Application

The Moon's craters are remnants of meteorite collisions. Suppose a fairly large asteroid that has a mass of 5.00×10^{12} kg (about a kilometer across) strikes the Moon at a speed of 15.0 km/s. (a) At what speed does the Moon recoil after the perfectly inelastic collision (the mass of the Moon is 7.36×10^{22} kg)? (b) How much kinetic energy is lost in the collision? Such an event may have been observed by medieval English monks who reported observing a red glow and subsequent haze about the Moon. (c) In October 2009, NASA crashed a rocket into the Moon, and analyzed the plume produced by the impact. (Significant amounts of water were detected.) Answer part (a) and (b) for this real-life experiment. The mass of the rocket was 2000 kg and its speed upon impact was 9000 km/h. How does the plume produced alter these results?

Solution:

(a) 1.02×10^{-6} m/s

(b) 5.63×10^{20} J (almost all KE lost)

(c) Recoil speed is 6.79×10^{-17} m/s, energy lost is 6.25×10^9 J. The plume will not affect the momentum result because the plume is still part of the Moon system. The plume may affect the kinetic energy result because a significant part of the initial kinetic energy may be transferred to the kinetic energy of the plume particles.

Exercise:

Problem: Professional Application

Two football players collide head-on in midair while trying to catch a thrown football. The first player is 95.0 kg and has an initial velocity of 6.00 m/s, while the second player is 115 kg and has an initial velocity of -3.50 m/s. What is their velocity just after impact if they cling together?

Exercise:**Problem:**

What is the speed of a garbage truck that is 1.20×10^4 kg and is initially moving at 25.0 m/s just after it hits and adheres to a trash can that is 80.0 kg and is initially at rest?

Solution:

24.8 m/s

Exercise:**Problem:**

During a circus act, an elderly performer thrills the crowd by catching a cannon ball shot at him. The cannon ball has a mass of 10.0 kg and the horizontal component of its velocity is 8.00 m/s when the 65.0-kg performer catches it. If the performer is on nearly frictionless roller skates, what is his recoil velocity?

Exercise:**Problem:**

(a) During an ice skating performance, an initially motionless 80.0-kg clown throws a fake barbell away. The clown's ice skates allow her to recoil frictionlessly. If the clown recoils with a velocity of 0.500 m/s and the barbell is thrown with a velocity of 10.0 m/s, what is the mass of the barbell? (b) How much kinetic energy is gained by this maneuver? (c) Where does the kinetic energy come from?

Solution:

(a) 4.00 kg

(b) 210 J

(c) The clown does work to throw the barbell, so the kinetic energy comes from the muscles of the clown. The muscles convert the chemical potential energy of ATP into kinetic energy.

Glossary

inelastic collision

a collision in which internal kinetic energy is not conserved

perfectly inelastic collision

a collision in which the colliding objects stick together

Introduction to Statics and Torque

class="introduction"

On a short time scale, rocks like these in Australia's Kings Canyon are static, or motionless relative to the Earth.

(credit:
freeaussiestock.com
)



What might desks, bridges, buildings, trees, and mountains have in common—at least in the eyes of a physicist? The answer is that they are ordinarily motionless relative to the Earth. Furthermore, their acceleration is zero because they remain motionless. That means they also have something in common with a car moving at a constant velocity, because anything with

a constant velocity also has an acceleration of zero. Now, the important part—Newton’s second law states that net $F = ma$, and so the net external force is zero for all stationary objects and for all objects moving at constant velocity. There are forces acting, but they are balanced. That is, they are in *equilibrium*.

Note:

Statics

Statics is the study of forces in equilibrium, a large group of situations that makes up a special case of Newton’s second law. We have already considered a few such situations; in this chapter, we cover the topic more thoroughly, including consideration of such possible effects as the rotation and deformation of an object by the forces acting on it.

How can we guarantee that a body is in equilibrium and what can we learn from systems that are in equilibrium? There are actually two conditions that must be satisfied to achieve equilibrium. These conditions are the topics of the first two sections of this chapter.

The First Condition for Equilibrium

- State the first condition of equilibrium.
- Explain static equilibrium.
- Explain dynamic equilibrium.

The first condition necessary to achieve equilibrium is the one already mentioned: the net external force on the system must be zero. Expressed as an equation, this is simply

Equation:

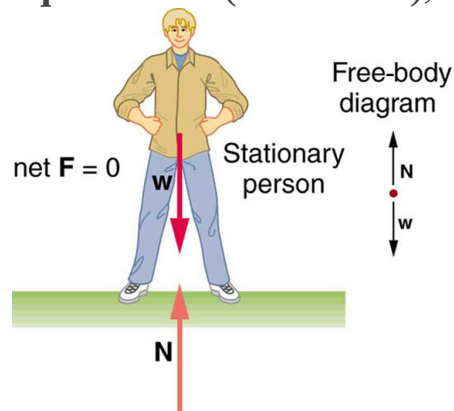
$$\text{net } \mathbf{F} = 0$$

Note that if net F is zero, then the net external force in *any* direction is zero. For example, the net external forces along the typical x- and y-axes are zero. This is written as

Equation:

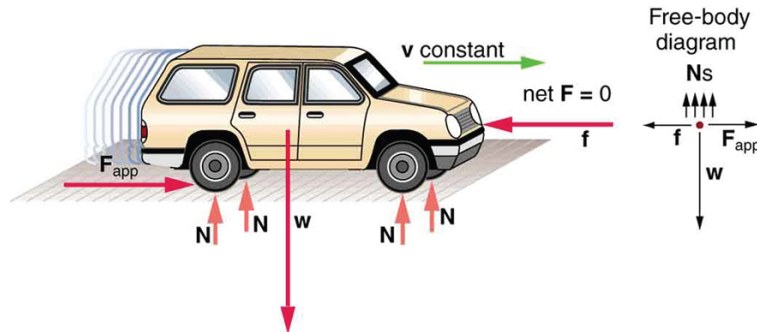
$$\text{net } F_x = 0 \text{ and } F_y = 0$$

[\[link\]](#) and [\[link\]](#) illustrate situations where net $F = 0$ for both **static equilibrium** (motionless), and **dynamic equilibrium** (constant velocity).



This motionless person is in static equilibrium. The forces acting on him add up to zero. Both

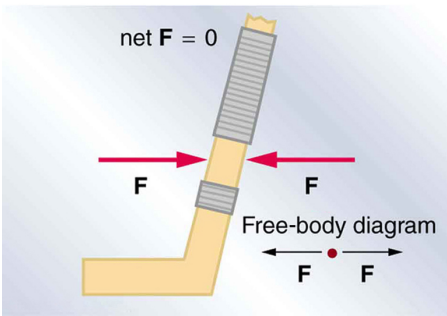
forces are vertical in this case.



This car is in dynamic equilibrium because it is moving at constant velocity. There are horizontal and vertical forces, but the net external force in any direction is zero. The applied force F_{app} between the tires and the road is balanced by air friction, and the weight of the car is supported by the normal forces, here shown to be equal for all four tires.

However, it is not sufficient for the net external force of a system to be zero for a system to be in equilibrium. Consider the two situations illustrated in [\[link\]](#) and [\[link\]](#) where forces are applied to an ice hockey stick lying flat on ice. The net external force is zero in both situations shown in the figure; but in one case, equilibrium is achieved, whereas in the other, it is not. In [\[link\]](#), the ice hockey stick remains motionless. But in [\[link\]](#), with the same forces applied in different places, the stick experiences accelerated rotation. Therefore, we know that the point at which a force is applied is another factor in determining whether or not equilibrium is achieved. This will be explored further in the next section.

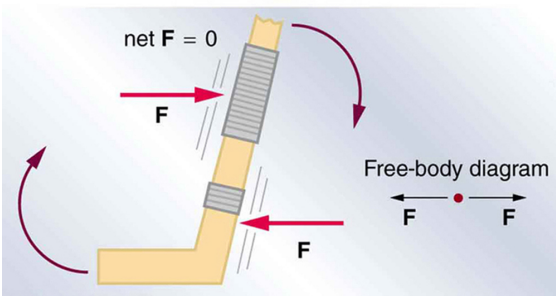
Equilibrium: remains stationary



An ice hockey stick lying flat on ice with two equal and opposite horizontal forces applied to it. Friction is negligible, and the gravitational force is balanced by the support of the ice (a normal force). Thus, net $F = 0$.

Equilibrium is achieved, which is static equilibrium in this case.

Nonequilibrium: rotation accelerates



The same forces are applied at other points and the stick

rotates—in fact, it experiences an accelerated rotation. Here net $F = 0$ but the system is *not* at equilibrium. Hence, the net $F = 0$ is a necessary—but not sufficient—condition for achieving equilibrium.

Note:

PhET Explorations: Torque

Investigate how torque causes an object to rotate. Discover the relationships between angular acceleration, moment of inertia, angular momentum and torque.

[Torqu
e](#)

Section Summary

- Statics is the study of forces in equilibrium.
- Two conditions must be met to achieve equilibrium, which is defined to be motion without linear or rotational acceleration.
- The first condition necessary to achieve equilibrium is that the net external force on the system must be zero, so that net $\mathbf{F} = 0$.

Conceptual Questions

Exercise:**Problem:**

What can you say about the velocity of a moving body that is in dynamic equilibrium? Draw a sketch of such a body using clearly labeled arrows to represent all external forces on the body.

Exercise:**Problem:**

Under what conditions can a rotating body be in equilibrium? Give an example.

Glossary

static equilibrium

a state of equilibrium in which the net external force and torque acting on a system is zero

dynamic equilibrium

a state of equilibrium in which the net external force and torque on a system moving with constant velocity are zero

The Second Condition for Equilibrium

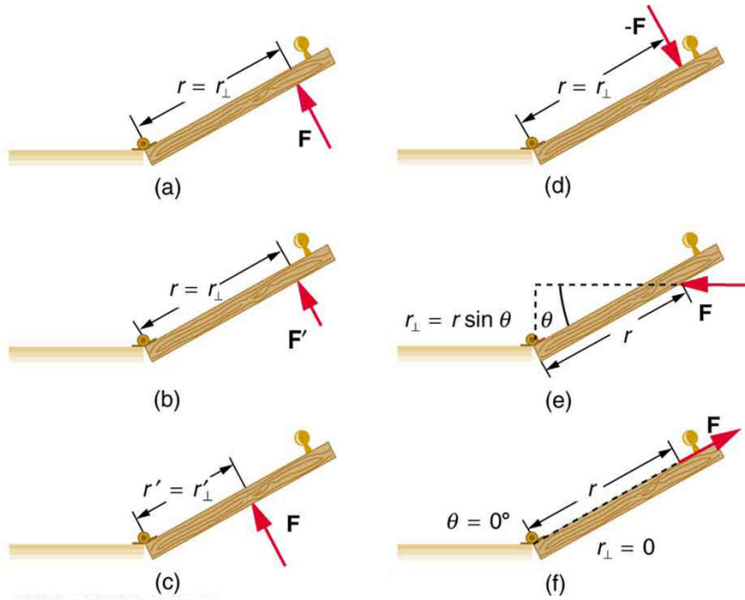
- State the second condition that is necessary to achieve equilibrium.
- Explain torque and the factors on which it depends.
- Describe the role of torque in rotational mechanics.

Note:

Torque

The second condition necessary to achieve equilibrium involves avoiding accelerated rotation (maintaining a constant angular velocity). A rotating body or system can be in equilibrium if its rate of rotation is constant and remains unchanged by the forces acting on it. To understand what factors affect rotation, let us think about what happens when you open an ordinary door by rotating it on its hinges.

Several familiar factors determine how effective you are in opening the door. See [\[link\]](#). First of all, the larger the force, the more effective it is in opening the door—obviously, the harder you push, the more rapidly the door opens. Also, the point at which you push is crucial. If you apply your force too close to the hinges, the door will open slowly, if at all. Most people have been embarrassed by making this mistake and bumping up against a door when it did not open as quickly as expected. Finally, the direction in which you push is also important. The most effective direction is perpendicular to the door—we push in this direction almost instinctively.



Torque is the turning or twisting effectiveness of a force, illustrated here for door rotation on its hinges (as viewed from overhead). Torque has both magnitude and direction. (a) Counterclockwise torque is produced by this force, which means that the door will rotate in a counterclockwise due to \mathbf{F} . Note that r_{\perp} is the perpendicular distance of the pivot from the line of action of the force. (b) A smaller counterclockwise torque is produced by a smaller force \mathbf{F}' acting at the same distance from the hinges (the pivot point). (c) The same force as in (a) produces a smaller counterclockwise torque when applied at a smaller distance from the hinges. (d) The same force as in (a), but acting in the opposite direction, produces a clockwise torque. (e) A smaller counterclockwise torque is produced by the same magnitude force acting at the same point

but in a different direction. Here, θ is less than 90° . (f) Torque is zero here since the force just pulls on the hinges, producing no rotation. In this case, $\theta = 0^\circ$.

The magnitude, direction, and point of application of the force are incorporated into the definition of the physical quantity called torque.

Torque is the rotational equivalent of a force. It is a measure of the effectiveness of a force in changing or accelerating a rotation (changing the angular velocity over a period of time). In equation form, the magnitude of torque is defined to be

Equation:

$$\tau = rF \sin \theta$$

where τ (the Greek letter tau) is the symbol for torque, r is the distance from the pivot point to the point where the force is applied, F is the magnitude of the force, and θ is the angle between the force and the vector directed from the point of application to the pivot point, as seen in [\[link\]](#) and [\[link\]](#). An alternative expression for torque is given in terms of the **perpendicular lever arm** r_\perp as shown in [\[link\]](#) and [\[link\]](#), which is defined as

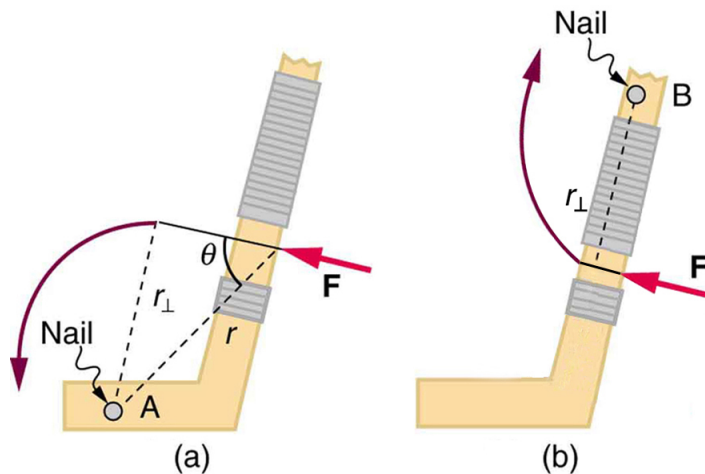
Equation:

$$r_\perp = r \sin \theta$$

so that

Equation:

$$\tau = r_\perp F.$$



A force applied to an object can produce a torque, which depends on the location of the pivot point. (a) The three factors r , F , and θ for pivot point A on a body are shown here— r is the distance from the chosen pivot point to the point where the force F is applied, and θ is the angle between \mathbf{F} and the vector directed from the point of application to the pivot point. If the object can rotate around point A, it will rotate counterclockwise. This means that torque is counterclockwise relative to pivot A. (b) In this case, point B is the pivot point. The torque from the applied force will cause a clockwise rotation around point B, and so it is a clockwise torque relative to B.

The perpendicular lever arm r_{\perp} is the shortest distance from the pivot point to the line along which \mathbf{F} acts; it is shown as a dashed line in [\[link\]](#) and [\[link\]](#). Note that the line segment that defines the distance r_{\perp} is perpendicular to \mathbf{F} , as its name implies. It is sometimes easier to find or

visualize r_{\perp} than to find both r and θ . In such cases, it may be more convenient to use $\tau = r_{\perp}F$ rather than $\tau = rF \sin \theta$ for torque, but both are equally valid.

The **SI unit of torque** is newtons times meters, usually written as $\text{N} \cdot \text{m}$. For example, if you push perpendicular to the door with a force of 40 N at a distance of 0.800 m from the hinges, you exert a torque of $32 \text{ N} \cdot \text{m}$ ($0.800 \text{ m} \times 40 \text{ N} \times \sin 90^\circ$) relative to the hinges. If you reduce the force to 20 N, the torque is reduced to $16 \text{ N} \cdot \text{m}$, and so on.

The torque is always calculated with reference to some chosen pivot point. For the same applied force, a different choice for the location of the pivot will give you a different value for the torque, since both r and θ depend on the location of the pivot. Any point in any object can be chosen to calculate the torque about that point. The object may not actually pivot about the chosen “pivot point.”

Note that for rotation in a plane, torque has two possible directions. Torque is either clockwise or counterclockwise relative to the chosen pivot point, as illustrated for points B and A, respectively, in [\[link\]](#). If the object can rotate about point A, it will rotate counterclockwise, which means that the torque for the force is shown as counterclockwise relative to A. But if the object can rotate about point B, it will rotate clockwise, which means the torque for the force shown is clockwise relative to B. Also, the magnitude of the torque is greater when the lever arm is longer.

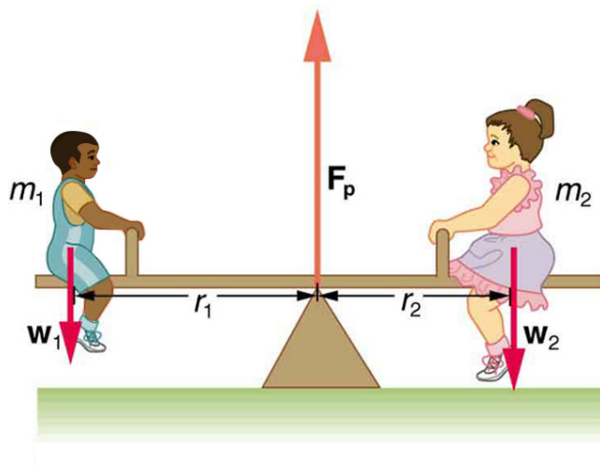
Now, *the second condition necessary to achieve equilibrium* is that *the net external torque on a system must be zero*. An external torque is one that is created by an external force. You can choose the point around which the torque is calculated. The point can be the physical pivot point of a system or any other point in space—but it must be the same point for all torques. If the second condition (net external torque on a system is zero) is satisfied for one choice of pivot point, it will also hold true for any other choice of pivot point in or out of the system of interest. (This is true only in an inertial frame of reference.) The second condition necessary to achieve equilibrium is stated in equation form as

Equation:

$$\text{net } \tau = 0$$

where net means total. Torques, which are in opposite directions are assigned opposite signs. A common convention is to call counterclockwise (ccw) torques positive and clockwise (cw) torques negative.

When two children balance a seesaw as shown in [\[link\]](#), they satisfy the two conditions for equilibrium. Most people have perfect intuition about seesaws, knowing that the lighter child must sit farther from the pivot and that a heavier child can keep a lighter one off the ground indefinitely.



Two children balancing a seesaw satisfy both conditions for equilibrium. The lighter child sits farther from the pivot to create a torque equal in magnitude to that of the heavier child.

Example:**She Saw Torques On A Seesaw**

The two children shown in [\[link\]](#) are balanced on a seesaw of negligible mass. (This assumption is made to keep the example simple—more

involved examples will follow.) The first child has a mass of 26.0 kg and sits 1.60 m from the pivot.(a) If the second child has a mass of 32.0 kg, how far is she from the pivot? (b) What is F_p , the supporting force exerted by the pivot?

Strategy

Both conditions for equilibrium must be satisfied. In part (a), we are asked for a distance; thus, the second condition (regarding torques) must be used, since the first (regarding only forces) has no distances in it. To apply the second condition for equilibrium, we first identify the system of interest to be the seesaw plus the two children. We take the supporting pivot to be the point about which the torques are calculated. We then identify all external forces acting on the system.

Solution (a)

The three external forces acting on the system are the weights of the two children and the supporting force of the pivot. Let us examine the torque produced by each. Torque is defined to be

Equation:

$$\tau = rF \sin \theta.$$

Here $\theta = 90^\circ$, so that $\sin \theta = 1$ for all three forces. That means $r_\perp = r$ for all three. The torques exerted by the three forces are first,

Equation:

$$\tau_1 = r_1 w_1$$

second,

Equation:

$$\tau_2 = -r_2 w_2$$

and third,

Equation:

$$\begin{aligned}\tau_p &= r_p F_p \\ &= 0 \cdot F_p \\ &= 0.\end{aligned}$$

Note that a minus sign has been inserted into the second equation because this torque is clockwise and is therefore negative by convention. Since F_p acts directly on the pivot point, the distance r_p is zero. A force acting on the pivot cannot cause a rotation, just as pushing directly on the hinges of a door will not cause it to rotate. Now, the second condition for equilibrium is that the sum of the torques on both children is zero. Therefore

Equation:

$$\tau_2 = -\tau_1,$$

or

Equation:

$$r_2 w_2 = r_1 w_1.$$

Weight is mass times the acceleration due to gravity. Entering mg for w , we get

Equation:

$$r_2 m_2 g = r_1 m_1 g.$$

Solve this for the unknown r_2 :

Equation:

$$r_2 = r_1 \frac{m_1}{m_2}.$$

The quantities on the right side of the equation are known; thus, r_2 is

Equation:

$$r_2 = (1.60 \text{ m}) \frac{26.0 \text{ kg}}{32.0 \text{ kg}} = 1.30 \text{ m}.$$

As expected, the heavier child must sit closer to the pivot (1.30 m versus 1.60 m) to balance the seesaw.

Solution (b)

This part asks for a force F_p . The easiest way to find it is to use the first condition for equilibrium, which is

Equation:

$$\text{net } \mathbf{F} = 0.$$

The forces are all vertical, so that we are dealing with a one-dimensional problem along the vertical axis; hence, the condition can be written as

Equation:

$$\text{net } F_y = 0$$

where we again call the vertical axis the y -axis. Choosing upward to be the positive direction, and using plus and minus signs to indicate the directions of the forces, we see that

Equation:

$$F_p - w_1 - w_2 = 0.$$

This equation yields what might have been guessed at the beginning:

Equation:

$$F_p = w_1 + w_2.$$

So, the pivot supplies a supporting force equal to the total weight of the system:

Equation:

$$F_p = m_1 g + m_2 g.$$

Entering known values gives

Equation:

$$\begin{aligned} F_p &= (26.0 \text{ kg})(9.80 \text{ m/s}^2) + (32.0 \text{ kg})(9.80 \text{ m/s}^2) \\ &= 568 \text{ N.} \end{aligned}$$

Discussion

The two results make intuitive sense. The heavier child sits closer to the pivot. The pivot supports the weight of the two children. Part (b) can also be solved using the second condition for equilibrium, since both distances are known, but only if the pivot point is chosen to be somewhere other than the location of the seesaw's actual pivot!

Several aspects of the preceding example have broad implications. First, the choice of the pivot as the point around which torques are calculated simplified the problem. Since F_p is exerted on the pivot point, its lever arm is zero. Hence, the torque exerted by the supporting force F_p is zero relative to that pivot point. The second condition for equilibrium holds for any choice of pivot point, and so we choose the pivot point to simplify the solution of the problem.

Second, the acceleration due to gravity canceled in this problem, and we were left with a ratio of masses. *This will not always be the case.* Always enter the correct forces—do not jump ahead to enter some ratio of masses.

Third, the weight of each child is distributed over an area of the seesaw, yet we treated the weights as if each force were exerted at a single point. This is not an approximation—the distances r_1 and r_2 are the distances to points directly below the **center of gravity** of each child. As we shall see in the next section, the mass and weight of a system can act as if they are located at a single point.

Finally, note that the concept of torque has an importance beyond static equilibrium. *Torque plays the same role in rotational motion that force plays in linear motion.* We will examine this in the next chapter.

Note:

Take-Home Experiment

Take a piece of modeling clay and put it on a table, then mash a cylinder down into it so that a ruler can balance on the round side of the cylinder while everything remains still. Put a penny 8 cm away from the pivot. Where would you need to put two pennies to balance? Three pennies?

Section Summary

- The second condition assures those torques are also balanced. Torque is the rotational equivalent of a force in producing a rotation and is

defined to be

Equation:

$$\tau = rF \sin \theta$$

where τ is torque, r is the distance from the pivot point to the point where the force is applied, F is the magnitude of the force, and θ is the angle between \mathbf{F} and the vector directed from the point where the force acts to the pivot point. The perpendicular lever arm r_{\perp} is defined to be

Equation:

$$r_{\perp} = r \sin \theta$$

so that

Equation:

$$\tau = r_{\perp} F.$$

- The perpendicular lever arm r_{\perp} is the shortest distance from the pivot point to the line along which F acts. The SI unit for torque is newton-meter (N·m). The second condition necessary to achieve equilibrium is that the net external torque on a system must be zero:

Equation:

$$\text{net } \tau = 0$$

By convention, counterclockwise torques are positive, and clockwise torques are negative.

Conceptual Questions

Exercise:

Problem:

What three factors affect the torque created by a force relative to a specific pivot point?

Exercise:**Problem:**

A wrecking ball is being used to knock down a building. One tall unsupported concrete wall remains standing. If the wrecking ball hits the wall near the top, is the wall more likely to fall over by rotating at its base or by falling straight down? Explain your answer. How is it most likely to fall if it is struck with the same force at its base? Note that this depends on how firmly the wall is attached at its base.

Exercise:**Problem:**

Mechanics sometimes put a length of pipe over the handle of a wrench when trying to remove a very tight bolt. How does this help? (It is also hazardous since it can break the bolt.)

Problems & Exercises**Exercise:****Problem:**

(a) When opening a door, you push on it perpendicularly with a force of 55.0 N at a distance of 0.850m from the hinges. What torque are you exerting relative to the hinges? (b) Does it matter if you push at the same height as the hinges?

Solution:

a) 46.8 N·m

b) It does not matter at what height you push. The torque depends on only the magnitude of the force applied and the perpendicular distance of the force's application from the hinges. (Children don't have a tougher time opening a door because they push lower than adults, they have a tougher time because they don't push far enough from the hinges.)

Exercise:

Problem:

When tightening a bolt, you push perpendicularly on a wrench with a force of 165 N at a distance of 0.140 m from the center of the bolt. (a) How much torque are you exerting in newton \times meters (relative to the center of the bolt)? (b) Convert this torque to footpounds.

Exercise:

Problem:

Two children push on opposite sides of a door during play. Both push horizontally and perpendicular to the door. One child pushes with a force of 17.5 N at a distance of 0.600 m from the hinges, and the second child pushes at a distance of 0.450 m. What force must the second child exert to keep the door from moving? Assume friction is negligible.

Solution:

23.3 N

Exercise:

Problem:

Use the second condition for equilibrium (net $\tau = 0$) to calculate F_p in [\[link\]](#), employing any data given or solved for in part (a) of the example.

Exercise:

Problem:

Repeat the seesaw problem in [\[link\]](#) with the center of mass of the seesaw 0.160 m to the left of the pivot (on the side of the lighter child) and assuming a mass of 12.0 kg for the seesaw. The other data given in the example remain unchanged. Explicitly show how you follow the steps in the Problem-Solving Strategy for static equilibrium.

Solution:

Given:

Equation:

$$\begin{aligned}m_1 &= 26.0 \text{ kg}, m_2 = 32.0 \text{ kg}, m_s = 12.0 \text{ kg}, \\r_1 &= 1.60 \text{ m}, r_s = 0.160 \text{ m}, \text{ find (a) } r_2, \text{ (b) } F_p\end{aligned}$$

a) Since children are balancing:

Equation:

$$\begin{aligned}\text{net } \tau_{\text{cw}} &= -\text{net } \tau_{\text{ccw}} \\ \Rightarrow w_1 r_1 + m_s g r_s &= w_2 r_2\end{aligned}$$

So, solving for r_2 gives:

Equation:

$$\begin{aligned}r_2 &= \frac{w_1 r_1 + m_s g r_s}{w_2} = \frac{m_1 g r_1 + m_s g r_s}{m_2 g} = \frac{m_1 r_1 + m_s r_s}{m_2} \\ &= \frac{(26.0 \text{ kg})(1.60 \text{ m}) + (12.0 \text{ kg})(0.160 \text{ m})}{32.0 \text{ kg}} \\ &= 1.36 \text{ m}\end{aligned}$$

b) Since the children are not moving:

Equation:

$$\text{net } F = 0 = F_p - w_1 - w_2 - w_s$$

$$\Rightarrow F_p = w_1 + w_2 + w_s$$

So that

Equation:

$$F_p = (26.0 \text{ kg} + 32.0 \text{ kg} + 12.0 \text{ kg})(9.80 \text{ m/s}^2)$$

$$= 686 \text{ N}$$

Glossary

torque

turning or twisting effectiveness of a force

perpendicular lever arm

the shortest distance from the pivot point to the line along which **F** lies

SI units of torque

newton times meters, usually written as N·m

center of gravity

the point where the total weight of the body is assumed to be concentrated

Applications of Statics, Including Problem-Solving Strategies

- Discuss the applications of Statics in real life.
- State and discuss various problem-solving strategies in Statics.

Statics can be applied to a variety of situations, ranging from raising a drawbridge to bad posture and back strain. We begin with a discussion of problem-solving strategies specifically used for statics. Since statics is a special case of Newton's laws, both the general problem-solving strategies and the special strategies for Newton's laws, discussed in [Problem-Solving Strategies](#), still apply.

Note:

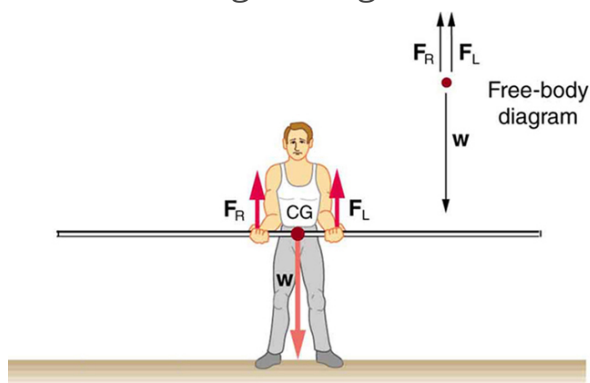
Problem-Solving Strategy: Static Equilibrium Situations

1. The first step is to determine whether or not the system is in **static equilibrium**. This condition is always the case when the *acceleration of the system is zero and accelerated rotation does not occur*.
2. It is particularly important to *draw a free body diagram for the system of interest*. Carefully label all forces, and note their relative magnitudes, directions, and points of application whenever these are known.
3. Solve the problem by applying either or both of the conditions for equilibrium (represented by the equations $\text{net } F = 0$ and $\text{net } \tau = 0$, depending on the list of known and unknown factors. If the second condition is involved, *choose the pivot point to simplify the solution*. Any pivot point can be chosen, but the most useful ones cause torques by unknown forces to be zero. (Torque is zero if the force is applied at the pivot (then $r = 0$), or along a line through the pivot point (then $\theta = 0$)). Always choose a convenient coordinate system for projecting forces.
4. *Check the solution to see if it is reasonable* by examining the magnitude, direction, and units of the answer. The importance of this last step never diminishes, although in unfamiliar applications, it is usually more difficult to judge reasonableness. These judgments become progressively easier with experience.

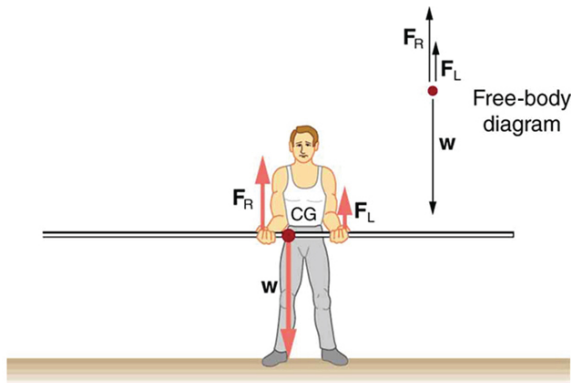
Now let us apply this problem-solving strategy for the pole vaulter shown in the three figures below. The pole is uniform and has a mass of 5.00 kg. In [\[link\]](#), the pole's cg lies halfway between the vaulter's hands. It seems reasonable that the force exerted by each hand is equal to half the weight of the pole, or 24.5 N. This obviously satisfies the first condition for equilibrium (net $F = 0$). The second condition (net $\tau = 0$) is also satisfied, as we can see by choosing the cg to be the pivot point. The weight exerts no torque about a pivot point located at the cg, since it is applied at that point and its lever arm is zero. The equal forces exerted by the hands are equidistant from the chosen pivot, and so they exert equal and opposite torques. Similar arguments hold for other systems where supporting forces are exerted symmetrically about the cg. For example, the four legs of a uniform table each support one-fourth of its weight.

In [\[link\]](#), a pole vaulter holding a pole with its cg halfway between his hands is shown. Each hand exerts a force equal to half the weight of the pole, $F_R = F_L = w/2$. (b) The pole vaulter moves the pole to his left, and the forces that the hands exert are no longer equal. See [\[link\]](#). If the pole is held with its cg to the left of the person, then he must push down with his right hand and up with his left. The forces he exerts are larger here because they are in opposite directions and the cg is at a long distance from either hand.

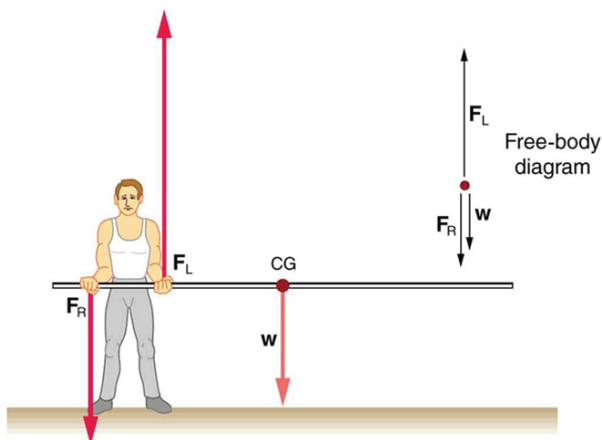
Similar observations can be made using a meter stick held at different locations along its length.



A pole vaulter holds a pole horizontally with both hands.



A pole vaulter is holding a pole horizontally with both hands. The center of gravity is near his right hand.



A pole vaulter is holding a pole horizontally with both hands. The center of gravity is to the left side of the vaulter.

If the pole vaulter holds the pole as shown in [\[link\]](#), the situation is not as simple. The total force he exerts is still equal to the weight of the pole, but it is not evenly divided between his hands. (If $F_L = F_R$, then the torques about the cg would not be equal since the lever arms are different.)

Logically, the right hand should support more weight, since it is closer to the cg. In fact, if the right hand is moved directly under the cg, it will support all the weight. This situation is exactly analogous to two people carrying a load; the one closer to the cg carries more of its weight. Finding the forces F_L and F_R is straightforward, as the next example shows.

If the pole vaulter holds the pole from near the end of the pole ([\[link\]](#)), the direction of the force applied by the right hand of the vaulter reverses its direction.

Example:

What Force Is Needed to Support a Weight Held Near Its CG?

For the situation shown in [\[link\]](#), calculate: (a) F_R , the force exerted by the right hand, and (b) F_L , the force exerted by the left hand. The hands are 0.900 m apart, and the cg of the pole is 0.600 m from the left hand.

Strategy

[\[link\]](#) includes a free body diagram for the pole, the system of interest. There is not enough information to use the first condition for equilibrium (net $F = 0$), since two of the three forces are unknown and the hand forces cannot be assumed to be equal in this case. There is enough information to use the second condition for equilibrium (net $\tau = 0$) if the pivot point is chosen to be at either hand, thereby making the torque from that hand zero. We choose to locate the pivot at the left hand in this part of the problem, to eliminate the torque from the left hand.

Solution for (a)

There are now only two nonzero torques, those from the gravitational force (τ_w) and from the push or pull of the right hand (τ_R). Stating the second condition in terms of clockwise and counterclockwise torques,

Equation:

$$\text{net } \tau_{\text{cw}} = -\text{net } \tau_{\text{ccw}}.$$

or the algebraic sum of the torques is zero.

Here this is

Equation:

$$\tau_R = -\tau_w$$

since the weight of the pole creates a counterclockwise torque and the right hand counters with a clockwise torque. Using the definition of torque, $\tau = rF \sin \theta$, noting that $\theta = 90^\circ$, and substituting known values, we obtain

Equation:

$$(0.900 \text{ m})(F_R) = (0.600 \text{ m})(mg).$$

Thus,

Equation:

$$\begin{aligned} F_R &= (0.667)(5.00 \text{ kg})(9.80 \text{ m/s}^2) \\ &= 32.7 \text{ N.} \end{aligned}$$

Solution for (b)

The first condition for equilibrium is based on the free body diagram in the figure. This implies that by Newton's second law:

Equation:

$$F_L + F_R - mg = 0$$

From this we can conclude:

Equation:

$$F_L + F_R = w = mg$$

Solving for F_L , we obtain

Equation:

$$\begin{aligned}
 F_L &= mg - F_R \\
 &= mg - 32.7 \text{ N} \\
 &= (5.00 \text{ kg})(9.80 \text{ m/s}^2) - 32.7 \text{ N} \\
 &= 16.3 \text{ N}
 \end{aligned}$$

Discussion

F_L is seen to be exactly half of F_R , as we might have guessed, since F_L is applied twice as far from the cg as F_R .

If the pole vaulter holds the pole as he might at the start of a run, shown in [\[link\]](#), the forces change again. Both are considerably greater, and one force reverses direction.

Note:

Take-Home Experiment

This is an experiment to perform while standing in a bus or a train. Stand facing sideways. How do you move your body to readjust the distribution of your mass as the bus accelerates and decelerates? Now stand facing forward. How do you move your body to readjust the distribution of your mass as the bus accelerates and decelerates? Why is it easier and safer to stand facing sideways rather than forward? Note: For your safety (and those around you), make sure you are holding onto something while you carry out this activity!

Note:

PhET Explorations: Balancing Act

Play with objects on a teeter totter to learn about balance. Test what you've learned by trying the Balance Challenge game.

https://phet.colorado.edu/sims/html/balancing-act/latest/balancing-act_en.html

Summary

- Statics can be applied to a variety of situations, ranging from raising a drawbridge to bad posture and back strain. We have discussed the problem-solving strategies specifically useful for statics. Statics is a special case of Newton's laws, both the general problem-solving strategies and the special strategies for Newton's laws, discussed in [Problem-Solving Strategies](#), still apply.

Conceptual Questions

Exercise:

Problem:

When visiting some countries, you may see a person balancing a load on the head. Explain why the center of mass of the load needs to be directly above the person's neck vertebrae.

Problems & Exercises

Exercise:

Problem:

To get up on the roof, a person (mass 70.0 kg) places a 6.00-m aluminum ladder (mass 10.0 kg) against the house on a concrete pad with the base of the ladder 2.00 m from the house. The ladder rests against a plastic rain gutter, which we can assume to be frictionless. The center of mass of the ladder is 2 m from the bottom. The person is standing 3 m from the bottom. What are the magnitudes of the forces on the ladder at the top and bottom?

Exercise:

Problem:

In [\[link\]](#), the cg of the pole held by the pole vaulter is 2.00 m from the left hand, and the hands are 0.700 m apart. Calculate the force exerted by (a) his right hand and (b) his left hand. (c) If each hand supports half the weight of the pole in [\[link\]](#), show that the second condition for equilibrium (net $\tau = 0$) is satisfied for a pivot other than the one located at the center of gravity of the pole. Explicitly show how you follow the steps in the Problem-Solving Strategy for static equilibrium described above.

Glossary

static equilibrium

equilibrium in which the acceleration of the system is zero and accelerated rotation does not occur

Simple Machines

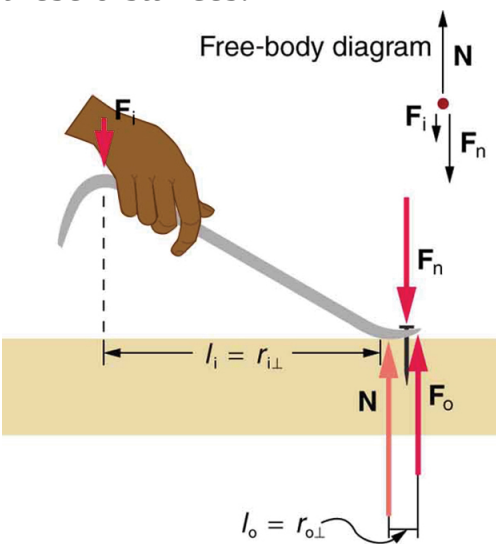
- Describe different simple machines.
- Calculate the mechanical advantage.

Simple machines are devices that can be used to multiply or augment a force that we apply – often at the expense of a distance through which we apply the force. The word for “machine” comes from the Greek word meaning “to help make things easier.” Levers, gears, pulleys, wedges, and screws are some examples of machines. Energy is still conserved for these devices because a machine cannot do more work than the energy put into it. However, machines can reduce the input force that is needed to perform the job. The ratio of output to input force magnitudes for any simple machine is called its **mechanical advantage** (MA).

Equation:

$$MA = \frac{F_o}{F_i}$$

One of the simplest machines is the lever, which is a rigid bar pivoted at a fixed place called the fulcrum. Torques are involved in levers, since there is rotation about a pivot point. Distances from the physical pivot of the lever are crucial, and we can obtain a useful expression for the MA in terms of these distances.



A nail puller is a lever with a large mechanical advantage. The external forces on the nail puller are represented by solid arrows. The force that the nail puller applies to the nail (\mathbf{F}_o) is not a force on the nail puller. The reaction force the nail exerts back on the puller (\mathbf{F}_n) is an external force and is equal and opposite to \mathbf{F}_o . The perpendicular lever arms of the input and output forces are l_i and l_o .

[\[link\]](#) shows a lever type that is used as a nail puller. Crowbars, seesaws, and other such levers are all analogous to this one. \mathbf{F}_i is the input force and \mathbf{F}_o is the output force. There are three vertical forces acting on the nail puller (the system of interest) – these are \mathbf{F}_i , \mathbf{F}_n , and \mathbf{N} . \mathbf{F}_n is the reaction force back on the system, equal and opposite to \mathbf{F}_o . (Note that \mathbf{F}_o is not a force on the system.) \mathbf{N} is the normal force upon the lever, and its torque is zero since it is exerted at the pivot. The torques due to \mathbf{F}_i and \mathbf{F}_n must be equal to each other if the nail is not moving, to satisfy the second condition for equilibrium (net $\tau = 0$). (In order for the nail to actually move, the torque due to \mathbf{F}_i must be ever-so-slightly greater than torque due to \mathbf{F}_n .) Hence,

Equation:

$$l_i F_i = l_o F_o$$

where l_i and l_o are the distances from where the input and output forces are applied to the pivot, as shown in the figure. Rearranging the last equation gives

Equation:

$$\frac{F_o}{F_i} = \frac{l_i}{l_o}.$$

What interests us most here is that the magnitude of the force exerted by the nail puller, F_o , is much greater than the magnitude of the input force applied to the puller at the other end, F_i . For the nail puller,

Equation:

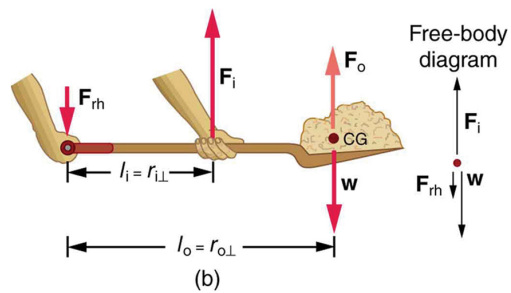
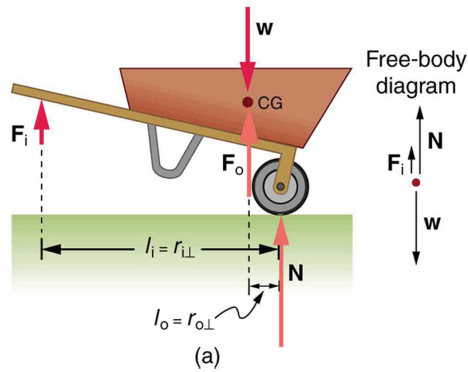
$$\text{MA} = \frac{F_o}{F_i} = \frac{l_i}{l_o}.$$

This equation is true for levers in general. For the nail puller, the MA is certainly greater than one. The longer the handle on the nail puller, the greater the force you can exert with it.

Two other types of levers that differ slightly from the nail puller are a wheelbarrow and a shovel, shown in [\[link\]](#). All these lever types are similar in that only three forces are involved – the input force, the output force, and the force on the pivot – and thus their MAs are given by $\text{MA} = \frac{F_o}{F_i}$ and $\text{MA} = \frac{d_1}{d_2}$, with distances being measured relative to the physical pivot.

The wheelbarrow and shovel differ from the nail puller because both the input and output forces are on the same side of the pivot.

In the case of the wheelbarrow, the output force or load is between the pivot (the wheel's axle) and the input or applied force. In the case of the shovel, the input force is between the pivot (at the end of the handle) and the load, but the input lever arm is shorter than the output lever arm. In this case, the MA is less than one.



(a) In the case of the wheelbarrow, the output force or load is between the pivot and the input force. The pivot is the wheel's axle. Here, the output force is greater than the input force.

Thus, a wheelbarrow enables you to lift much heavier loads than you could with your body alone. (b) In the case of the shovel, the input force is between the pivot and the load, but the input lever arm is shorter than the output lever arm. The pivot is at the handle held by the right hand. Here, the output force

(supporting the shovel's load) is less than the input force (from the hand nearest the load), because the input is exerted closer to the pivot than is the output.

Example:**What is the Advantage for the Wheelbarrow?**

In the wheelbarrow of [\[link\]](#), the load has a perpendicular lever arm of 7.50 cm, while the hands have a perpendicular lever arm of 1.02 m. (a) What upward force must you exert to support the wheelbarrow and its load if their combined mass is 45.0 kg? (b) What force does the wheelbarrow exert on the ground?

Strategy

Here, we use the concept of mechanical advantage.

Solution

(a) In this case, $\frac{F_o}{F_i} = \frac{l_i}{l_o}$ becomes

Equation:

$$F_i = F_o \frac{l_o}{l_i}.$$

Adding values into this equation yields

Equation:

$$F_i = (45.0 \text{ kg}) \left(9.80 \text{ m/s}^2 \right) \frac{0.075 \text{ m}}{1.02 \text{ m}} = 32.4 \text{ N}.$$

The free-body diagram (see [\[link\]](#)) gives the following normal force:

$F_i + N = W$. Therefore, $N = (45.0 \text{ kg}) \left(9.80 \text{ m/s}^2 \right) - 32.4 \text{ N} = 409 \text{ N}$

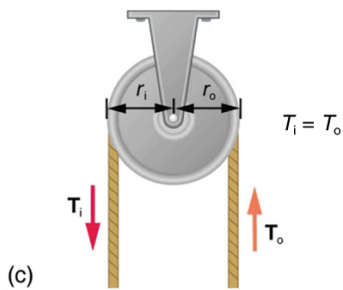
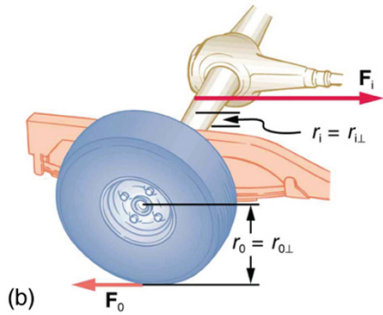
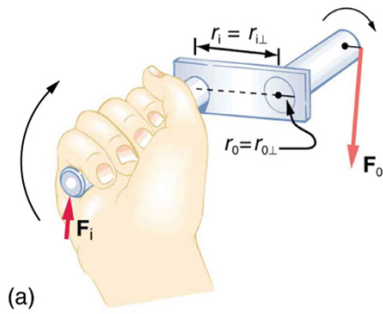
. N is the normal force acting on the wheel; by Newton's third law, the force the wheel exerts on the ground is 409 N.

Discussion

An even longer handle would reduce the force needed to lift the load. The MA here is $MA = 1.02/0.0750 = 13.6$.

Another very simple machine is the inclined plane. Pushing a cart up a plane is easier than lifting the same cart straight up to the top using a ladder, because the applied force is less. However, the work done in both cases (assuming the work done by friction is negligible) is the same. Inclined lanes or ramps were probably used during the construction of the Egyptian pyramids to move large blocks of stone to the top.

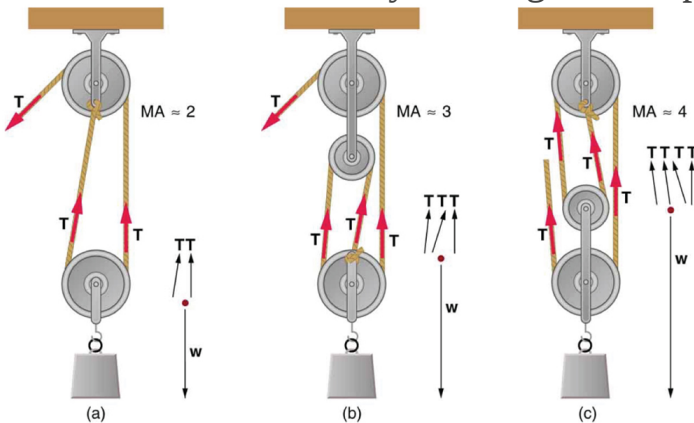
A crank is a lever that can be rotated 360° about its pivot, as shown in [\[link\]](#). Such a machine may not look like a lever, but the physics of its actions remain the same. The MA for a crank is simply the ratio of the radii r_i/r_o . Wheels and gears have this simple expression for their MAs too. The MA can be greater than 1, as it is for the crank, or less than 1, as it is for the simplified car axle driving the wheels, as shown. If the axle's radius is 2.0 cm and the wheel's radius is 24.0 cm, then $MA = 2.0/24.0 = 0.083$ and the axle would have to exert a force of 12,000 N on the wheel to enable it to exert a force of 1000 N on the ground.



(a) A crank is a type of lever that can be rotated 360° about its pivot. Cranks are usually designed to have a large MA. (b) A simplified automobile axle drives a wheel, which has a much larger diameter than the axle. The MA is less than 1.

(c) An ordinary pulley is used to lift a heavy load. The pulley changes the direction of the force T exerted by the cord without changing its magnitude. Hence, this machine has an MA of 1.

An ordinary pulley has an MA of 1; it only changes the direction of the force and not its magnitude. Combinations of pulleys, such as those illustrated in [\[link\]](#), are used to multiply force. If the pulleys are friction-free, then the force output is approximately an integral multiple of the tension in the cable. The number of cables pulling directly upward on the system of interest, as illustrated in the figures given below, is approximately the MA of the pulley system. Since each attachment applies an external force in approximately the same direction as the others, they add, producing a total force that is nearly an integral multiple of the input force T .



(a) The combination of pulleys is used to multiply force. The force is an integral multiple of tension if the pulleys are frictionless. This pulley

system has two cables attached to its load, thus applying a force of approximately $2T$. This machine has $MA \approx 2$. (b) Three pulleys are used to lift a load in such a way that the mechanical advantage is about 3. Effectively, there are three cables attached to the load. (c) This pulley system applies a force of $4T$, so that it has $MA \approx 4$. Effectively, four cables are pulling on the system of interest.

Section Summary

- Simple machines are devices that can be used to multiply or augment a force that we apply – often at the expense of a distance through which we have to apply the force.
- The ratio of output to input forces for any simple machine is called its mechanical advantage
- A few simple machines are the lever, nail puller, wheelbarrow, crank, etc.

Conceptual Questions

Exercise:

Problem:

Scissors are like a double-lever system. Which of the simple machines in [\[link\]](#) and [\[link\]](#) is most analogous to scissors?

Exercise:

Problem:

Suppose you pull a nail at a constant rate using a nail puller as shown in [\[link\]](#). Is the nail puller in equilibrium? What if you pull the nail with some acceleration – is the nail puller in equilibrium then? In which case is the force applied to the nail puller larger and why?

Exercise:**Problem:**

Why are the forces exerted on the outside world by the limbs of our bodies usually much smaller than the forces exerted by muscles inside the body?

Exercise:**Problem:**

Explain why the forces in our joints are several times larger than the forces we exert on the outside world with our limbs. Can these forces be even greater than muscle forces (see previous Question)?

Problems & Exercises**Exercise:****Problem:**

What is the mechanical advantage of a nail puller—similar to the one shown in [\[link\]](#)—where you exert a force 45 cm from the pivot and the nail is 1.8 cm on the other side? What minimum force must you exert to apply a force of 1250 N to the nail?

Solution:

25

50 N

Exercise:**Problem:**

Suppose you needed to raise a 250-kg mower a distance of 6.0 cm above the ground to change a tire. If you had a 2.0-m long lever, where would you place the fulcrum if your force was limited to 300 N?

Exercise:**Problem:**

a) What is the mechanical advantage of a wheelbarrow, such as the one in [\[link\]](#), if the center of gravity of the wheelbarrow and its load has a perpendicular lever arm of 5.50 cm, while the hands have a perpendicular lever arm of 1.02 m? (b) What upward force should you exert to support the wheelbarrow and its load if their combined mass is 55.0 kg? (c) What force does the wheel exert on the ground?

Solution:

- a) $MA = 18.5$
- b) $F_i = 29.1 \text{ N}$
- c) 510 N downward

Exercise:**Problem:**

A typical car has an axle with 1.10 cm radius driving a tire with a radius of 27.5 cm. What is its mechanical advantage assuming the very simplified model in [\[link\]](#)(b)?

Exercise:**Problem:**

What force does the nail puller in [\[link\]](#) exert on the supporting surface? The nail puller has a mass of 2.10 kg.

Solution:

$$1.3 \times 10^3 \text{ N}$$

Exercise:**Problem:**

If you used an ideal pulley of the type shown in [\[link\]](#)(a) to support a car engine of mass 115 kg, (a) What would be the tension in the rope? (b) What force must the ceiling supply, assuming you pull straight down on the rope? Neglect the pulley system's mass.

Exercise:**Problem:**

Repeat [\[link\]](#) for the pulley shown in [\[link\]](#)(c), assuming you pull straight up on the rope. The pulley system's mass is 7.00 kg.

Solution:

a) $T = 299 \text{ N}$

b) 897 N upward

Glossary

mechanical advantage

the ratio of output to input forces for any simple machine

Introduction to Temperature, Kinetic Theory, and the Gas Laws

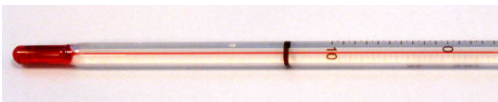
class="introduction"

The welder's
gloves and
helmet
protect him
from the
electric arc
that transfers
enough
thermal
energy to
melt the rod,
spray sparks,
and burn the
retina of an
unprotected
eye. The
thermal
energy can
be felt on
exposed skin
a few meters
away, and its
light can be
seen for
kilometers.
(credit:
Kevin S.
O'Brien/U.S
. Navy)



Heat is something familiar to each of us. We feel the warmth of the summer Sun, the chill of a clear summer night, the heat of coffee after a winter stroll, and the cooling effect of our sweat. Heat transfer is maintained by temperature differences. Manifestations of **heat transfer**—the movement of heat energy from one place or material to another—are apparent throughout the universe. Heat from beneath Earth's surface is brought to the surface in flows of incandescent lava. The Sun warms Earth's surface and is the source of much of the energy we find on it. Rising levels of atmospheric carbon dioxide threaten to trap more of the Sun's energy, perhaps fundamentally altering the ecosphere. In space, supernovas explode, briefly radiating more heat than an entire galaxy does.

What is heat? How do we define it? How is it related to temperature? What are heat's effects? How is it related to other forms of energy and to work? We will find that, in spite of the richness of the phenomena, there is a small set of underlying physical principles that unite the subjects and tie them to other fields.



In a typical thermometer like this one, the alcohol, with a red dye, expands

more rapidly than the glass containing it. When the thermometer's temperature increases, the liquid from the bulb is forced into the narrow tube, producing a large change in the length of the column for a small change in temperature.

(credit: Chemical Engineer, Wikimedia Commons)

Temperature

- Define temperature.
- Convert temperatures between the Celsius, Fahrenheit, and Kelvin scales.
- Define thermal equilibrium.
- State the zeroth law of thermodynamics.

The concept of temperature has evolved from the common concepts of hot and cold. Human perception of what feels hot or cold is a relative one. For example, if you place one hand in hot water and the other in cold water, and then place both hands in tepid water, the tepid water will feel cool to the hand that was in hot water, and warm to the one that was in cold water. The scientific definition of temperature is less ambiguous than your senses of hot and cold. **Temperature** is operationally defined to be what we measure with a thermometer. (Many physical quantities are defined solely in terms of how they are measured. We shall see later how temperature is related to the kinetic energies of atoms and molecules, a more physical explanation.) Two accurate thermometers, one placed in hot water and the other in cold water, will show the hot water to have a higher temperature. If they are then placed in the tepid water, both will give identical readings (within measurement uncertainties). In this section, we discuss temperature, its measurement by thermometers, and its relationship to thermal equilibrium. Again, temperature is the quantity measured by a thermometer.

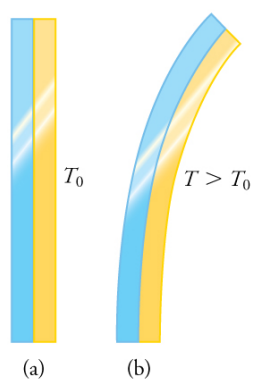
Note:

Misconception Alert: Human Perception vs. Reality

On a cold winter morning, the wood on a porch feels warmer than the metal of your bike. The wood and bicycle are in thermal equilibrium with the outside air, and are thus the same temperature. They *feel* different because of the difference in the way that they conduct heat away from your skin. The metal conducts heat away from your body faster than the wood does (see more about conductivity in [Conduction](#)). This is just one example demonstrating that the human sense of hot and cold is not determined by temperature alone.

Another factor that affects our perception of temperature is humidity. Most people feel much hotter on hot, humid days than on hot, dry days. This is because on humid days, sweat does not evaporate from the skin as efficiently as it does on dry days. It is the evaporation of sweat (or water from a sprinkler or pool) that cools us off.

Any physical property that depends on temperature, and whose response to temperature is reproducible, can be used as the basis of a thermometer. Because many physical properties depend on temperature, the variety of thermometers is remarkable. For example, volume increases with temperature for most substances. This property is the basis for the common alcohol thermometer, the old mercury thermometer, and the bimetallic strip ([link](#)). Other properties used to measure temperature include electrical resistance and color, as shown in [link](#), and the emission of infrared radiation, as shown in [link](#).



The curvature of a bimetallic strip depends on

temperature.
 (a) The strip is straight at the starting temperature, where its two components have the same length.

(b) At a higher temperature, this strip bends to the right, because the metal on the left has expanded more than the metal on the right.



Each of the six squares on this plastic (liquid crystal)

thermometer contains a film of a different heat-sensitive liquid crystal material. Below 95°F, all six squares are black.

When the plastic thermometer is exposed to temperature that increases to 95°F, the first liquid crystal square changes color. When the temperature increases above 96.8°F the second liquid crystal square also changes color, and so forth. (credit: Arkrishna, Wikimedia Commons)



Fireman Jason Ormand uses a pyrometer to check the temperature of an aircraft carrier's ventilation system. Infrared radiation (whose emission varies with temperature)

from the vent is measured and a temperature readout is quickly produced.

Infrared measurements are also frequently used as a measure of body temperature.

These modern thermometers, placed in the ear canal, are more accurate than alcohol thermometers placed under the tongue or in the armpit. (credit: Lamel J. Hinton/U.S. Navy)

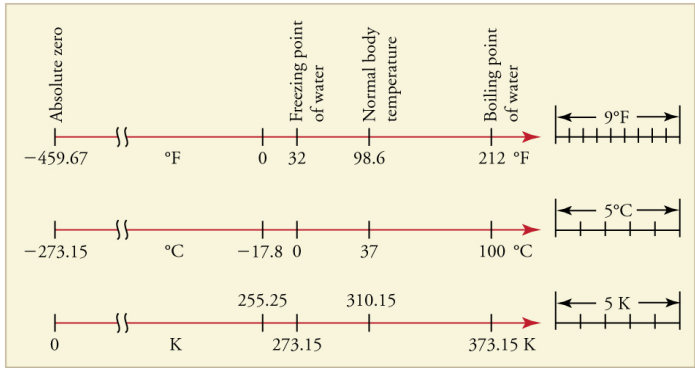
Temperature Scales

Thermometers are used to measure temperature according to well-defined scales of measurement, which use pre-defined reference points to help compare quantities. The three most common temperature scales are the Fahrenheit, Celsius, and Kelvin scales. A temperature scale can be created by identifying two easily reproducible temperatures. The freezing and boiling temperatures of water at standard atmospheric pressure are commonly used.

The **Celsius** scale (which replaced the slightly different *centigrade* scale) has the freezing point of water at 0°C and the boiling point at 100°C . Its unit is the **degree Celsius** ($^{\circ}\text{C}$). On the **Fahrenheit** scale (still the most frequently used in the United States), the freezing point of water is at 32°F and the boiling point is at 212°F . The unit of temperature on this scale is the **degree Fahrenheit** ($^{\circ}\text{F}$). Note that a temperature difference of one degree Celsius is greater than a temperature difference of one degree Fahrenheit. Only 100 Celsius degrees

span the same range as 180 Fahrenheit degrees, thus one degree on the Celsius scale is 1.8 times larger than one degree on the Fahrenheit scale $180/100 = 9/5$.

The **Kelvin** scale is the temperature scale that is commonly used in science. It is an *absolute temperature* scale defined to have 0 K at the lowest possible temperature, called **absolute zero**. The official temperature unit on this scale is the *kelvin*, which is abbreviated K, and is not accompanied by a degree sign. The freezing and boiling points of water are 273.15 K and 373.15 K, respectively. Thus, the magnitude of temperature differences is the same in units of kelvins and degrees Celsius. Unlike other temperature scales, the Kelvin scale is an absolute scale. It is used extensively in scientific work because a number of physical quantities, such as the volume of an ideal gas, are directly related to absolute temperature. The kelvin is the SI unit used in scientific work.



Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales, rounded to the nearest degree. The relative sizes of the scales are also shown.

The relationships between the three common temperature scales is shown in [\[link\]](#). Temperatures on these scales can be converted using the equations in [\[link\]](#).

To convert from ...	Use this equation ...	Also written as ...

To convert from ...	Use this equation ...	Also written as ...
Celsius to Fahrenheit	$T(^{\circ}\text{F}) = \frac{9}{5}T(^{\circ}\text{C}) + 32$	$T_{\text{F}} = \frac{9}{5}T_{\text{C}} + 32$
Fahrenheit to Celsius	$T(^{\circ}\text{C}) = \frac{5}{9}(T(^{\circ}\text{F}) - 32)$	$T_{\text{C}} = \frac{5}{9}(T_{\text{F}} - 32)$
Celsius to Kelvin	$T(\text{K}) = T(^{\circ}\text{C}) + 273.15$	$T_{\text{K}} = T_{\text{C}} + 273.15$
Kelvin to Celsius	$T(^{\circ}\text{C}) = T(\text{K}) - 273.15$	$T_{\text{C}} = T_{\text{K}} - 273.15$
Fahrenheit to Kelvin	$T(\text{K}) = \frac{5}{9}(T(^{\circ}\text{F}) - 32) + 273.15$	$T_{\text{K}} = \frac{5}{9}(T_{\text{F}} - 32) + 273.15$
Kelvin to Fahrenheit	$T(^{\circ}\text{F}) = \frac{9}{5}(T(\text{K}) - 273.15) + 32$	$T_{\text{F}} = \frac{9}{5}(T_{\text{K}} - 273.15) + 32$

Temperature Conversions

Notice that the conversions between Fahrenheit and Kelvin look quite complicated. In fact, they are simple combinations of the conversions between Fahrenheit and Celsius, and the conversions between Celsius and Kelvin.

Example:

Converting between Temperature Scales: Room Temperature

“Room temperature” is generally defined to be 25°C . (a) What is room temperature in $^{\circ}\text{F}$?

(b) What is it in K?

Strategy

To answer these questions, all we need to do is choose the correct conversion equations and plug in the known values.

Solution for (a)

1. Choose the right equation. To convert from °C to °F, use the equation

Equation:

$$T_{\text{°F}} = \frac{9}{5}T_{\text{°C}} + 32.$$

2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{°F}} = \frac{9}{5}25^{\circ}\text{C} + 32 = 77^{\circ}\text{F}.$$

Solution for (b)

1. Choose the right equation. To convert from °C to K, use the equation

Equation:

$$T_{\text{K}} = T_{\text{°C}} + 273.15.$$

2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{K}} = 25^{\circ}\text{C} + 273.15 = 298 \text{ K}.$$

Example:**Converting between Temperature Scales: the Reaumur Scale**

The Reaumur scale is a temperature scale that was used widely in Europe in the 18th and 19th centuries. On the Reaumur temperature scale, the freezing point of water is 0°R and the boiling temperature is 80°R. If “room temperature” is 25°C on the Celsius scale, what is it on the Reaumur scale?

Strategy

To answer this question, we must compare the Reaumur scale to the Celsius scale. The difference between the freezing point and boiling point of water on the Reaumur scale is 80°R. On the Celsius scale it is 100°C. Therefore $100^{\circ}\text{C} = 80^{\circ}\text{R}$. Both scales start at 0° for freezing, so we can derive a simple formula to convert between temperatures on the two scales.

Solution

1. Derive a formula to convert from one scale to the other:

Equation:

$$T_{\text{°R}} = \frac{0.8^{\circ}\text{R}}{^{\circ}\text{C}} \times T_{\text{°C}}.$$

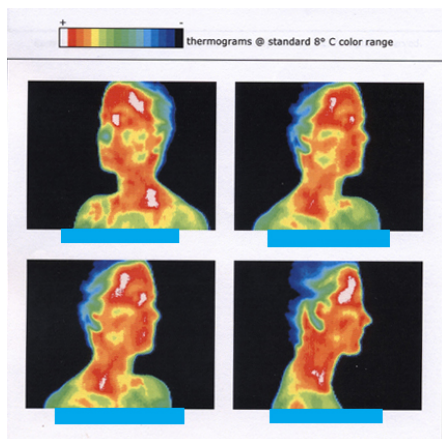
2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{R}} = \frac{0.8^{\circ}\text{R}}{^{\circ}\text{C}} \times 25^{\circ}\text{C} = 20^{\circ}\text{R}.$$

Temperature Ranges in the Universe

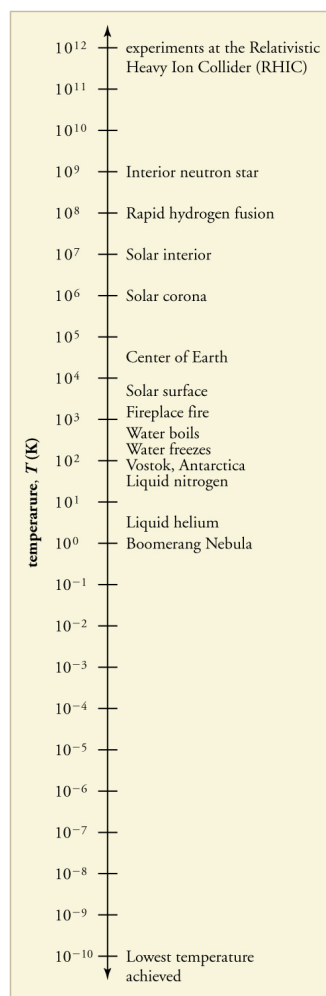
[\[link\]](#) shows the wide range of temperatures found in the universe. Human beings have been known to survive with body temperatures within a small range, from 24°C to 44°C (75°F to 111°F). The average normal body temperature is usually given as 37.0°C (98.6°F), and variations in this temperature can indicate a medical condition: a fever, an infection, a tumor, or circulatory problems (see [\[link\]](#)).



This image of radiation from a person's body (an infrared thermograph) shows the location of temperature abnormalities in the upper body. Dark blue corresponds to cold areas and red to white corresponds to hot areas. An elevated temperature might be an indication of malignant tissue (a cancerous tumor in the breast, for example), while a depressed temperature

might be due to a decline in blood flow from a clot. In this case, the abnormalities are caused by a condition called hyperhidrosis.
(credit: Porcelina81, Wikimedia Commons)

The lowest temperatures ever recorded have been measured during laboratory experiments: 4.5×10^{-10} K at the Massachusetts Institute of Technology (USA), and 1.0×10^{-10} K at Helsinki University of Technology (Finland). In comparison, the coldest recorded place on Earth's surface is Vostok, Antarctica at 183 K (-89°C), and the coldest place (outside the lab) known in the universe is the Boomerang Nebula, with a temperature of 1 K.

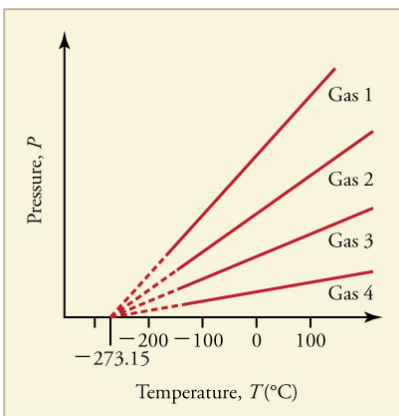


Each increment on this logarithmic scale indicates an increase by a factor of ten, and thus illustrates the tremendous range of temperatures in nature. Note that zero on a logarithmic scale would occur off the bottom of the page at infinity.

Note:

Making Connections: Absolute Zero

What is absolute zero? Absolute zero is the temperature at which all molecular motion has ceased. The concept of absolute zero arises from the behavior of gases. [\[link\]](#) shows how the pressure of gases at a constant volume decreases as temperature decreases. Various scientists have noted that the pressures of gases extrapolate to zero at the same temperature, -273.15°C . This extrapolation implies that there is a lowest temperature. This temperature is called *absolute zero*. Today we know that most gases first liquefy and then freeze, and it is not actually possible to reach absolute zero. The numerical value of absolute zero temperature is -273.15°C or 0 K.



Graph of pressure versus temperature for various

gases kept at a constant volume. Note that all of the graphs extrapolate to zero pressure at the same temperature.

Thermal Equilibrium and the Zeroth Law of Thermodynamics

Thermometers actually take their *own* temperature, not the temperature of the object they are measuring. This raises the question of how we can be certain that a thermometer measures the temperature of the object with which it is in contact. It is based on the fact that any two systems placed in *thermal contact* (meaning heat transfer can occur between them) will reach the same temperature. That is, heat will flow from the hotter object to the cooler one until they have exactly the same temperature. The objects are then in **thermal equilibrium**, and no further changes will occur. The systems interact and change because their temperatures differ, and the changes stop once their temperatures are the same. Thus, if enough time is allowed for this transfer of heat to run its course, the temperature a thermometer registers *does* represent the system with which it is in thermal equilibrium. Thermal equilibrium is established when two bodies are in contact with each other and can freely exchange energy.

Furthermore, experimentation has shown that if two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system C, then A is also in thermal equilibrium with C. This conclusion may seem obvious, because all three have the same temperature, but it is basic to thermodynamics. It is called the **zeroth law of thermodynamics**.

Note:

The Zeroth Law of Thermodynamics

If two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

This law was postulated in the 1930s, after the first and second laws of thermodynamics had been developed and named. It is called the *zeroth law* because it comes logically before the first and second laws (discussed in [Thermodynamics](#)). An example of this law in action is seen in babies in incubators: babies in incubators normally have very few clothes on, so to an observer they look as if they may not be warm enough. However, the temperature of the air, the cot, and the baby is the same, because they are in thermal equilibrium, which is accomplished by maintaining air temperature to keep the baby comfortable.

Exercise:
Check Your Understanding

Problem: Does the temperature of a body depend on its size?

Solution:

No, the system can be divided into smaller parts each of which is at the same temperature. We say that the temperature is an *intensive* quantity. Intensive quantities are independent of size.

Section Summary

- Temperature is the quantity measured by a thermometer.
- Temperature is related to the average kinetic energy of atoms and molecules in a system.
- Absolute zero is the temperature at which there is no molecular motion.
- There are three main temperature scales: Celsius, Fahrenheit, and Kelvin.
- Temperatures on one scale can be converted to temperatures on another scale using the following equations:

Equation:

$$T_{\text{°F}} = \frac{9}{5}T_{\text{°C}} + 32$$

Equation:

$$T_{\text{°C}} = \frac{5}{9}(T_{\text{°F}} - 32)$$

Equation:

$$T_{\text{K}} = T_{\text{°C}} + 273.15$$

Equation:

$$T_{\text{°C}} = T_{\text{K}} - 273.15$$

- Systems are in thermal equilibrium when they have the same temperature.
- Thermal equilibrium occurs when two bodies are in contact with each other and can freely exchange energy.
- The zeroth law of thermodynamics states that when two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

Conceptual Questions

Exercise:

Problem: What does it mean to say that two systems are in thermal equilibrium?

Exercise:

Problem:

Give an example of a physical property that varies with temperature and describe how it is used to measure temperature.

Exercise:

Problem:

When a cold alcohol thermometer is placed in a hot liquid, the column of alcohol goes *down* slightly before going up. Explain why.

Exercise:

Problem:

If you add boiling water to a cup at room temperature, what would you expect the final equilibrium temperature of the unit to be? You will need to include the surroundings as part of the system. Consider the zeroth law of thermodynamics.

Problems & Exercises

Exercise:

Problem: What is the Fahrenheit temperature of a person with a 39.0°C fever?

Solution:

102°F

Exercise:

Problem:

Frost damage to most plants occurs at temperatures of 28.0°F or lower. What is this temperature on the Kelvin scale?

Exercise:

Problem:

To conserve energy, room temperatures are kept at 68.0°F in the winter and 78.0°F in the summer. What are these temperatures on the Celsius scale?

Solution:

20.0°C and 25.6°C

Exercise:**Problem:**

A tungsten light bulb filament may operate at 2900 K . What is its Fahrenheit temperature? What is this on the Celsius scale?

Exercise:**Problem:**

The surface temperature of the Sun is about 5750 K . What is this temperature on the Fahrenheit scale?

Solution:

9890°F

Exercise:**Problem:**

One of the hottest temperatures ever recorded on the surface of Earth was 134°F in Death Valley, CA. What is this temperature in Celsius degrees? What is this temperature in Kelvin?

Exercise:**Problem:**

(a) Suppose a cold front blows into your locale and drops the temperature by 40.0 Fahrenheit degrees. How many degrees Celsius does the temperature decrease when there is a 40.0°F decrease in temperature? (b) Show that any change in temperature in Fahrenheit degrees is nine-fifths the change in Celsius degrees.

Solution:

(a) 22.2°C

$$\begin{aligned}
 \Delta T(^{\circ}\text{F}) &= T_2(^{\circ}\text{F}) - T_1(^{\circ}\text{F}) \\
 \text{(b)} \quad &= \frac{9}{5}T_2(^{\circ}\text{C}) + 32.0^{\circ} - \left(\frac{9}{5}T_1(^{\circ}\text{C}) + 32.0^{\circ}\right) \\
 &= \frac{9}{5}(T_2(^{\circ}\text{C}) - T_1(^{\circ}\text{C})) = \frac{9}{5}\Delta T(^{\circ}\text{C})
 \end{aligned}$$

Exercise:

Problem:

(a) At what temperature do the Fahrenheit and Celsius scales have the same numerical value? (b) At what temperature do the Fahrenheit and Kelvin scales have the same numerical value?

Glossary

temperature

the quantity measured by a thermometer

Celsius scale

temperature scale in which the freezing point of water is 0°C and the boiling point of water is 100°C

degree Celsius

unit on the Celsius temperature scale

Fahrenheit scale

temperature scale in which the freezing point of water is 32°F and the boiling point of water is 212°F

degree Fahrenheit

unit on the Fahrenheit temperature scale

Kelvin scale

temperature scale in which 0 K is the lowest possible temperature, representing absolute zero

absolute zero

the lowest possible temperature; the temperature at which all molecular motion ceases

thermal equilibrium

the condition in which heat no longer flows between two objects that are in contact; the two objects have the same temperature

zeroth law of thermodynamics

law that states that if two objects are in thermal equilibrium, and a third object is in thermal equilibrium with one of those objects, it is also in thermal equilibrium with the other object

Thermal Expansion of Solids and Liquids

- Define and describe thermal expansion.
- Calculate the linear expansion of an object given its initial length, change in temperature, and coefficient of linear expansion.
- Calculate the volume expansion of an object given its initial volume, change in temperature, and coefficient of volume expansion.
- Calculate thermal stress on an object given its original volume, temperature change, volume change, and bulk modulus.



Thermal expansion joints like these in the Auckland Harbour Bridge in New Zealand allow bridges to change length without buckling. (credit: Ingolfson, Wikimedia Commons)

The expansion of alcohol in a thermometer is one of many commonly encountered examples of **thermal expansion**, the change in size or volume of a given mass with temperature. Hot air rises because its volume increases, which causes the hot air's density to be smaller than the density of surrounding air, causing a buoyant (upward) force on the hot air. The same happens in all liquids and gases, driving natural heat transfer upwards in homes, oceans, and weather systems. Solids also undergo thermal expansion. Railroad tracks and bridges, for example, have expansion joints to allow them to freely expand and contract with temperature changes.

What are the basic properties of thermal expansion? First, thermal expansion is clearly related to temperature change. The greater the temperature change, the more a bimetallic strip will bend. Second, it depends on the material. In a thermometer, for example, the expansion of alcohol is much greater than the expansion of the glass containing it.

What is the underlying cause of thermal expansion? As is discussed in [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#), an increase in temperature implies an increase in the kinetic energy of the individual atoms. In a solid, unlike in a gas, the atoms or molecules are closely packed together, but their kinetic energy (in the form of small, rapid vibrations) pushes neighboring atoms or molecules apart from each other. This neighbor-to-neighbor pushing results in a slightly greater distance, on average, between neighbors, and adds up to a larger size for the whole body. For most substances under ordinary conditions, there is no preferred direction, and an increase in temperature will increase the solid's size by a certain fraction in each dimension.

Note:

Linear Thermal Expansion—Thermal Expansion in One Dimension

The change in length ΔL is proportional to length L . The dependence of thermal expansion on temperature, substance, and length is summarized in the equation

Equation:

$$\Delta L = \alpha L \Delta T,$$

where ΔL is the change in length L , ΔT is the change in temperature, and α is the **coefficient of linear expansion**, which varies slightly with temperature.

[\[link\]](#) lists representative values of the coefficient of linear expansion, which may have units of $1/^{\circ}\text{C}$ or $1/\text{K}$. Because the size of a kelvin and a degree Celsius are the same, both α and ΔT can be expressed in units of kelvins or degrees Celsius. The equation $\Delta L = \alpha L \Delta T$ is accurate for small changes in temperature and can be used for large changes in temperature if an average value of α is used.

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Solids		
Aluminum	25×10^{-6}	75×10^{-6}
Brass	19×10^{-6}	56×10^{-6}
Copper	17×10^{-6}	51×10^{-6}

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Gold	14×10^{-6}	42×10^{-6}
Iron or Steel	12×10^{-6}	35×10^{-6}
Invar (Nickel-iron alloy)	0.9×10^{-6}	2.7×10^{-6}
Lead	29×10^{-6}	87×10^{-6}
Silver	18×10^{-6}	54×10^{-6}
Glass (ordinary)	9×10^{-6}	27×10^{-6}
Glass (Pyrex®)	3×10^{-6}	9×10^{-6}
Quartz	0.4×10^{-6}	1×10^{-6}

Material	Coefficient of linear expansion	Coefficient of volume expansion
	$\alpha(1/^{\circ}\text{C})$	$\beta(1/^{\circ}\text{C})$
Concrete, Brick	$\sim 12 \times 10^{-6}$	$\sim 36 \times 10^{-6}$
Marble (average)	7×10^{-6}	2.1×10^{-5}
Liquids		
Ether		1650×10^{-6}
Ethyl alcohol		1100×10^{-6}
Petrol		950×10^{-6}
Glycerin		500×10^{-6}
Mercury		180×10^{-6}

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Water		210×10^{-6}
Gases		
Air and most other gases at atmospheric pressure		3400×10^{-6}

Thermal Expansion Coefficients at 20°C[\[footnote\]](#)

Values for liquids and gases are approximate.

Example:

Calculating Linear Thermal Expansion: The Golden Gate Bridge

The main span of San Francisco's Golden Gate Bridge is 1275 m long at its coldest. The bridge is exposed to temperatures ranging from -15°C to 40°C . What is its change in length between these temperatures? Assume that the bridge is made entirely of steel.

Strategy

Use the equation for linear thermal expansion $\Delta L = \alpha L \Delta T$ to calculate the change in length, ΔL . Use the coefficient of linear expansion, α , for steel from [\[link\]](#), and note that the change in temperature, ΔT , is 55°C .

Solution

Plug all of the known values into the equation to solve for ΔL .

Equation:

$$\Delta L = \alpha L \Delta T = \left(\frac{12 \times 10^{-6}}{^{\circ}\text{C}} \right) (1275 \text{ m})(55^{\circ}\text{C}) = 0.84 \text{ m}.$$

Discussion

Although not large compared with the length of the bridge, this change in length is observable. It is generally spread over many expansion joints so that the expansion at each joint is small.

Thermal Expansion in Two and Three Dimensions

Objects expand in all dimensions, as illustrated in [\[link\]](#). That is, their areas and volumes, as well as their lengths, increase with temperature. Holes also get larger with temperature. If you cut a hole in a metal plate, the remaining material will expand exactly as it would if the plug was still in place. The plug would get bigger, and so the hole must get bigger too. (Think of the ring of neighboring atoms or molecules on the wall of the hole as pushing each other farther apart as temperature increases. Obviously, the ring of neighbors must get slightly larger, so the hole gets slightly larger).

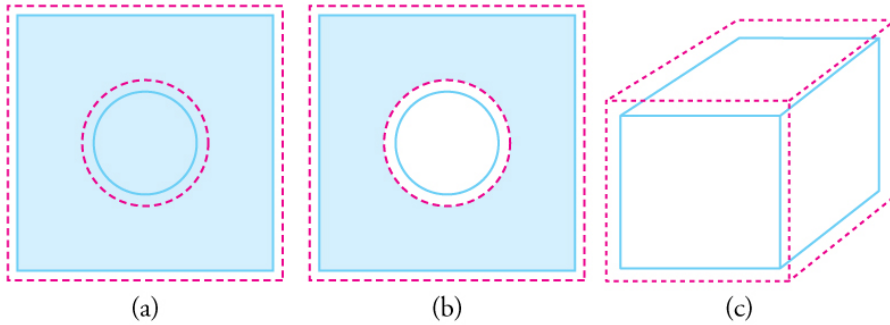
Note:**Thermal Expansion in Two Dimensions**

For small temperature changes, the change in area ΔA is given by

Equation:

$$\Delta A = 2\alpha A \Delta T,$$

where ΔA is the change in area A , ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.



In general, objects expand in all directions as temperature increases. In these drawings, the original boundaries of the objects are shown with solid lines, and the expanded boundaries with dashed lines. (a) Area increases because both length and width increase. The area of a circular plug also increases. (b) If the plug is removed, the hole it leaves becomes larger with increasing temperature, just as if the expanding plug were still in place. (c) Volume also increases, because all three dimensions increase.

Note:

Thermal Expansion in Three Dimensions

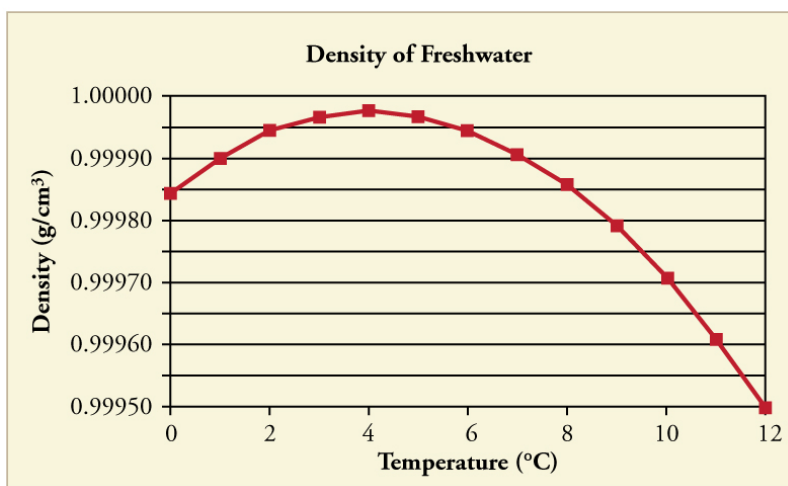
The change in volume ΔV is very nearly $\Delta V = 3\alpha V \Delta T$. This equation is usually written as

Equation:

$$\Delta V = \beta V \Delta T,$$

where β is the **coefficient of volume expansion** and $\beta \approx 3\alpha$. Note that the values of β in [\[link\]](#) are almost exactly equal to 3α .

In general, objects will expand with increasing temperature. Water is the most important exception to this rule. Water expands with increasing temperature (its density *decreases*) when it is at temperatures greater than 4°C (40°F). However, it expands with *decreasing* temperature when it is between $+4^{\circ}\text{C}$ and 0°C (40°F to 32°F). Water is densest at $+4^{\circ}\text{C}$. (See [\[link\]](#).) Perhaps the most striking effect of this phenomenon is the freezing of water in a pond. When water near the surface cools down to 4°C it is denser than the remaining water and thus will sink to the bottom. This “turnover” results in a layer of warmer water near the surface, which is then cooled. Eventually the pond has a uniform temperature of 4°C . If the temperature in the surface layer drops below 4°C , the water is less dense than the water below, and thus stays near the top. As a result, the pond surface can completely freeze over. The ice on top of liquid water provides an insulating layer from winter’s harsh exterior air temperatures. Fish and other aquatic life can survive in 4°C water beneath ice, due to this unusual characteristic of water. It also produces circulation of water in the pond that is necessary for a healthy ecosystem of the body of water.



The density of water as a function of temperature. Note that the thermal expansion is actually very small. The maximum density at $+4^{\circ}\text{C}$ is only 0.0075% greater than the density at 2°C , and 0.012% greater than that at 0°C .

Note:**Making Connections: Real-World Connections—Filling the Tank**

Differences in the thermal expansion of materials can lead to interesting effects at the gas station. One example is the dripping of gasoline from a freshly filled tank on a hot day. Gasoline starts out at the temperature of the ground under the gas station, which is cooler than the air temperature above. The gasoline cools the steel tank when it is filled. Both gasoline and steel tank expand as they warm to air temperature, but gasoline expands much more than steel, and so it may overflow.

This difference in expansion can also cause problems when interpreting the gasoline gauge. The actual amount (mass) of gasoline left in the tank when the gauge hits “empty” is a lot less in the summer than in the winter. The gasoline has the same volume as it does in the winter when the “add fuel” light goes on, but because the gasoline has expanded, there is less mass. If you are used to getting another 40 miles on “empty” in the winter, beware—you will probably run out much more quickly in the summer.



Because the gas expands more than the gas tank with increasing temperature, you can't drive as many miles on “empty” in the summer as you can in the winter.

(credit: Hector Alejandro,
Flickr)

Example:

Calculating Thermal Expansion: Gas vs. Gas Tank

Suppose your 60.0-L (15.9-gal) steel gasoline tank is full of gas, so both the tank and the gasoline have a temperature of 15.0°C. How much gasoline has spilled by the time they warm to 35.0°C?

Strategy

The tank and gasoline increase in volume, but the gasoline increases more, so the amount spilled is the difference in their volume changes. (The gasoline tank can be treated as solid steel.) We can use the equation for volume expansion to calculate the change in volume of the gasoline and of the tank.

Solution

1. Use the equation for volume expansion to calculate the increase in volume of the steel tank:

Equation:

$$\Delta V_s = \beta_s V_s \Delta T.$$

2. The increase in volume of the gasoline is given by this equation:

Equation:

$$\Delta V_{\text{gas}} = \beta_{\text{gas}} V_{\text{gas}} \Delta T.$$

3. Find the difference in volume to determine the amount spilled as

Equation:

$$V_{\text{spill}} = \Delta V_{\text{gas}} - \Delta V_s.$$

Alternatively, we can combine these three equations into a single equation. (Note that the original volumes are equal.)

Equation:

$$\begin{aligned}
 V_{\text{spill}} &= (\beta_{\text{gas}} - \beta_{\text{s}})V\Delta T \\
 &= [(950 - 35) \times 10^{-6} / ^\circ\text{C}] (60.0 \text{ L})(20.0^\circ\text{C}) \\
 &= 1.10 \text{ L.}
 \end{aligned}$$

Discussion

This amount is significant, particularly for a 60.0-L tank. The effect is so striking because the gasoline and steel expand quickly. The rate of change in thermal properties is discussed in [Heat and Heat Transfer Methods](#).

If you try to cap the tank tightly to prevent overflow, you will find that it leaks anyway, either around the cap or by bursting the tank. Tightly constricting the expanding gas is equivalent to compressing it, and both liquids and solids resist being compressed with extremely large forces. To avoid rupturing rigid containers, these containers have air gaps, which allow them to expand and contract without stressing them.

Thermal Stress

Thermal stress is created by thermal expansion or contraction (see [Elasticity: Stress and Strain](#) for a discussion of stress and strain). Thermal stress can be destructive, such as when expanding gasoline ruptures a tank. It can also be useful, for example, when two parts are joined together by heating one in manufacturing, then slipping it over the other and allowing the combination to cool. Thermal stress can explain many phenomena, such as the weathering of rocks and pavement by the expansion of ice when it freezes.

Example:

Calculating Thermal Stress: Gas Pressure

What pressure would be created in the gasoline tank considered in [\[link\]](#), if the gasoline increases in temperature from 15.0°C to 35.0°C without being allowed to expand? Assume that the bulk modulus B for gasoline is $1.00 \times 10^9 \text{ N/m}^2$. (For more on bulk modulus, see [Elasticity: Stress and Strain](#).)

Strategy

To solve this problem, we must use the following equation, which relates a change in volume ΔV to pressure:

Equation:

$$\Delta V = \frac{1}{B} \frac{F}{A} V_0,$$

where F/A is pressure, V_0 is the original volume, and B is the bulk modulus of the material involved. We will use the amount spilled in [\[link\]](#) as the change in volume, ΔV .

Solution

1. Rearrange the equation for calculating pressure:

Equation:

$$P = \frac{F}{A} = \frac{\Delta V}{V_0} B.$$

2. Insert the known values. The bulk modulus for gasoline is $B = 1.00 \times 10^9 \text{ N/m}^2$. In the previous example, the change in volume $\Delta V = 1.10 \text{ L}$ is the amount that would spill. Here, $V_0 = 60.0 \text{ L}$ is the original volume of the gasoline. Substituting these values into the equation, we obtain

Equation:

$$P = \frac{1.10 \text{ L}}{60.0 \text{ L}} (1.00 \times 10^9 \text{ Pa}) = 1.83 \times 10^7 \text{ Pa}.$$

Discussion

This pressure is about 2500 lb/in^2 , *much* more than a gasoline tank can handle.

Forces and pressures created by thermal stress are typically as great as that in the example above. Railroad tracks and roadways can buckle on hot days if they lack sufficient expansion joints. (See [\[link\]](#).) Power lines sag more in the summer than in the winter, and will snap in cold weather if there is

insufficient slack. Cracks open and close in plaster walls as a house warms and cools. Glass cooking pans will crack if cooled rapidly or unevenly, because of differential contraction and the stresses it creates. (Pyrex® is less susceptible because of its small coefficient of thermal expansion.) Nuclear reactor pressure vessels are threatened by overly rapid cooling, and although none have failed, several have been cooled faster than considered desirable. Biological cells are ruptured when foods are frozen, detracting from their taste. Repeated thawing and freezing accentuate the damage. Even the oceans can be affected. A significant portion of the rise in sea level that is resulting from global warming is due to the thermal expansion of sea water.



Thermal stress contributes to the formation of potholes.
(credit: Editor5807, Wikimedia Commons)

Metal is regularly used in the human body for hip and knee implants. Most implants need to be replaced over time because, among other things, metal does not bond with bone. Researchers are trying to find better metal coatings that would allow metal-to-bone bonding. One challenge is to find a coating that has an expansion coefficient similar to that of metal. If the

expansion coefficients are too different, the thermal stresses during the manufacturing process lead to cracks at the coating-metal interface.

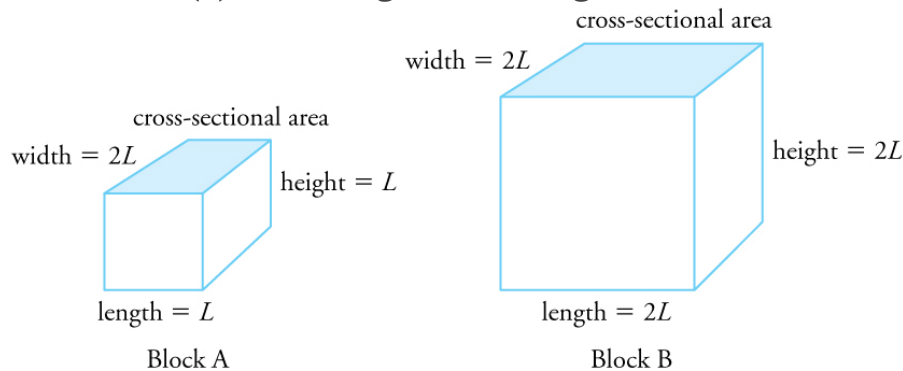
Another example of thermal stress is found in the mouth. Dental fillings can expand differently from tooth enamel. It can give pain when eating ice cream or having a hot drink. Cracks might occur in the filling. Metal fillings (gold, silver, etc.) are being replaced by composite fillings (porcelain), which have smaller coefficients of expansion, and are closer to those of teeth.

Exercise:

Check Your Understanding

Problem:

Two blocks, A and B, are made of the same material. Block A has dimensions $l \times w \times h = L \times 2L \times L$ and Block B has dimensions $2L \times 2L \times 2L$. If the temperature changes, what is (a) the change in the volume of the two blocks, (b) the change in the cross-sectional area $l \times w$, and (c) the change in the height h of the two blocks?



Solution:

(a) The change in volume is proportional to the original volume. Block A has a volume of $L \times 2L \times L = 2L^3$. Block B has a volume of $2L \times 2L \times 2L = 8L^3$, which is 4 times that of Block A. Thus the change in volume of Block B should be 4 times the change in volume of Block A.

(b) The change in area is proportional to the area. The cross-sectional area of Block A is $L \times 2L = 2L^2$, while that of Block B is

$2L \times 2L = 4L^2$. Because cross-sectional area of Block B is twice that of Block A, the change in the cross-sectional area of Block B is twice that of Block A.

(c) The change in height is proportional to the original height. Because the original height of Block B is twice that of A, the change in the height of Block B is twice that of Block A.

Section Summary

- Thermal expansion is the increase, or decrease, of the size (length, area, or volume) of a body due to a change in temperature.
- Thermal expansion is large for gases, and relatively small, but not negligible, for liquids and solids.
- Linear thermal expansion is

Equation:

$$\Delta L = \alpha L \Delta T,$$

where ΔL is the change in length L , ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.

- The change in area due to thermal expansion is

Equation:

$$\Delta A = 2\alpha A \Delta T,$$

where ΔA is the change in area.

- The change in volume due to thermal expansion is

Equation:

$$\Delta V = \beta V \Delta T,$$

where β is the coefficient of volume expansion and $\beta \approx 3\alpha$. Thermal stress is created when thermal expansion is constrained.

Conceptual Questions

Exercise:**Problem:**

Thermal stresses caused by uneven cooling can easily break glass cookware. Explain why Pyrex®, a glass with a small coefficient of linear expansion, is less susceptible.

Exercise:**Problem:**

Water expands significantly when it freezes: a volume increase of about 9% occurs. As a result of this expansion and because of the formation and growth of crystals as water freezes, anywhere from 10% to 30% of biological cells are burst when animal or plant material is frozen. Discuss the implications of this cell damage for the prospect of preserving human bodies by freezing so that they can be thawed at some future date when it is hoped that all diseases are curable.

Exercise:**Problem:**

One method of getting a tight fit, say of a metal peg in a hole in a metal block, is to manufacture the peg slightly larger than the hole. The peg is then inserted when at a different temperature than the block. Should the block be hotter or colder than the peg during insertion? Explain your answer.

Exercise:**Problem:**

Does it really help to run hot water over a tight metal lid on a glass jar before trying to open it? Explain your answer.

Exercise:

Problem:

Liquids and solids expand with increasing temperature, because the kinetic energy of a body's atoms and molecules increases. Explain why some materials *shrink* with increasing temperature.

Problems & Exercises**Exercise:****Problem:**

The height of the Washington Monument is measured to be 170 m on a day when the temperature is 35.0°C . What will its height be on a day when the temperature falls to -10.0°C ? Although the monument is made of limestone, assume that its thermal coefficient of expansion is the same as marble's.

Solution:

169.98 m

Exercise:**Problem:**

How much taller does the Eiffel Tower become at the end of a day when the temperature has increased by 15°C ? Its original height is 321 m and you can assume it is made of steel.

Exercise:**Problem:**

What is the change in length of a 3.00-cm-long column of mercury if its temperature changes from 37.0°C to 40.0°C , assuming the mercury is unconstrained?

Solution:

$$5.4 \times 10^{-6} \text{ m}$$

Exercise:**Problem:**

How large an expansion gap should be left between steel railroad rails if they may reach a maximum temperature 35.0°C greater than when they were laid? Their original length is 10.0 m.

Exercise:**Problem:**

You are looking to purchase a small piece of land in Hong Kong. The price is “only” \$60,000 per square meter! The land title says the dimensions are $20 \text{ m} \times 30 \text{ m}$. By how much would the total price change if you measured the parcel with a steel tape measure on a day when the temperature was 20°C above normal?

Solution:

Because the area gets smaller, the price of the land DECREASES by ~\$17,000.

Exercise:**Problem:**

Global warming will produce rising sea levels partly due to melting ice caps but also due to the expansion of water as average ocean temperatures rise. To get some idea of the size of this effect, calculate the change in length of a column of water 1.00 km high for a temperature increase of 1.00°C . Note that this calculation is only approximate because ocean warming is not uniform with depth.

Exercise:**Problem:**

Show that 60.0 L of gasoline originally at 15.0°C will expand to 61.1 L when it warms to 35.0°C , as claimed in [\[link\]](#).

Solution:

Equation:

$$\begin{aligned} V &= V_0 + \Delta V = V_0(1 + \beta\Delta T) \\ &= (60.00 \text{ L})[1 + (950 \times 10^{-6}/^{\circ}\text{C})(35.0^{\circ}\text{C} - 15.0^{\circ}\text{C})] \\ &= 61.1 \text{ L} \end{aligned}$$

Exercise:

Problem:

(a) Suppose a meter stick made of steel and one made of invar (an alloy of iron and nickel) are the same length at 0°C . What is their difference in length at 22.0°C ? (b) Repeat the calculation for two 30.0-m-long surveyor's tapes.

Exercise:

Problem:

(a) If a 500-mL glass beaker is filled to the brim with ethyl alcohol at a temperature of 5.00°C , how much will overflow when its temperature reaches 22.0°C ? (b) How much less water would overflow under the same conditions?

Solution:

(a) 9.35 mL

(b) 7.56 mL

Exercise:

Problem:

Most automobiles have a coolant reservoir to catch radiator fluid that may overflow when the engine is hot. A radiator is made of copper and is filled to its 16.0-L capacity when at 10.0°C. What volume of radiator fluid will overflow when the radiator and fluid reach their 95.0°C operating temperature, given that the fluid's volume coefficient of expansion is $\beta = 400 \times 10^{-6} / ^\circ\text{C}$? Note that this coefficient is approximate, because most car radiators have operating temperatures of greater than 95.0°C.

Exercise:**Problem:**

A physicist makes a cup of instant coffee and notices that, as the coffee cools, its level drops 3.00 mm in the glass cup. Show that this decrease cannot be due to thermal contraction by calculating the decrease in level if the 350 cm³ of coffee is in a 7.00-cm-diameter cup and decreases in temperature from 95.0°C to 45.0°C. (Most of the drop in level is actually due to escaping bubbles of air.)

Solution:

0.832 mm

Exercise:**Problem:**

(a) The density of water at 0°C is very nearly 1000 kg/m³ (it is actually 999.84 kg/m³), whereas the density of ice at 0°C is 917 kg/m³. Calculate the pressure necessary to keep ice from expanding when it freezes, neglecting the effect such a large pressure would have on the freezing temperature. (This problem gives you only an indication of how large the forces associated with freezing water might be.) (b) What are the implications of this result for biological cells that are frozen?

Exercise:**Problem:**

Show that $\beta \approx 3\alpha$, by calculating the change in volume ΔV of a cube with sides of length L .

Solution:

We know how the length changes with temperature: $\Delta L = \alpha L_0 \Delta T$. Also we know that the volume of a cube is related to its length by $V = L^3$, so the final volume is then $V = V_0 + \Delta V = (L_0 + \Delta L)^3$. Substituting for ΔL gives

Equation:

$$V = (L_0 + \alpha L_0 \Delta T)^3 = L_0^3 (1 + \alpha \Delta T)^3.$$

Now, because $\alpha \Delta T$ is small, we can use the binomial expansion:

Equation:

$$V \approx L_0^3 (1 + 3\alpha \Delta T) = L_0^3 + 3\alpha L_0^3 \Delta T.$$

So writing the length terms in terms of volumes gives $V = V_0 + \Delta V \approx V_0 + 3\alpha V_0 \Delta T$, and so

Equation:

$$\Delta V = \beta V_0 \Delta T \approx 3\alpha V_0 \Delta T, \text{ or } \beta \approx 3\alpha.$$

Glossary

thermal expansion

the change in size or volume of an object with change in temperature

coefficient of linear expansion

α , the change in length, per unit length, per 1°C change in temperature; a constant used in the calculation of linear expansion; the coefficient of linear expansion depends on the material and to some degree on the temperature of the material

coefficient of volume expansion

β , the change in volume, per unit volume, per 1°C change in temperature

thermal stress

stress caused by thermal expansion or contraction

The Ideal Gas Law

- State the ideal gas law in terms of molecules and in terms of moles.
- Use the ideal gas law to calculate pressure change, temperature change, volume change, or the number of molecules or moles in a given volume.
- Use Avogadro's number to convert between number of molecules and number of moles.

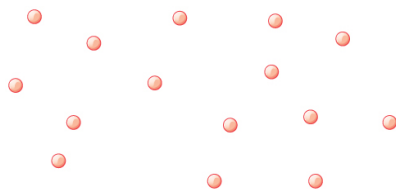


The air inside this hot air balloon flying over Putrajaya, Malaysia, is hotter than the ambient air. As a result, the balloon experiences a buoyant force pushing it upward.
(credit: Kevin Poh, Flickr)

In this section, we continue to explore the thermal behavior of gases. In particular, we examine the characteristics of atoms and molecules that compose gases. (Most gases, for example nitrogen, N_2 , and oxygen, O_2 , are composed of two or more atoms. We will primarily use the term “molecule” in discussing a gas because the term can also be applied to monatomic gases, such as helium.)

Gases are easily compressed. We can see evidence of this in [\[link\]](#), where you will note that gases have the *largest* coefficients of volume expansion. The large coefficients mean that gases expand and contract very rapidly with temperature changes. In addition, you will note that most gases expand at the *same* rate, or have the same β . This raises the question as to why gases should all act in nearly the same way, when liquids and solids have widely varying expansion rates.

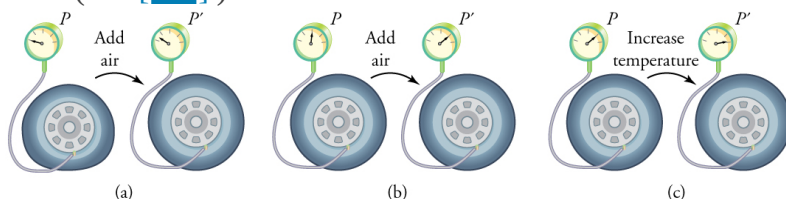
The answer lies in the large separation of atoms and molecules in gases, compared to their sizes, as illustrated in [\[link\]](#). Because atoms and molecules have large separations, forces between them can be ignored, except when they collide with each other during collisions. The motion of atoms and molecules (at temperatures well above the boiling temperature) is fast, such that the gas occupies all of the accessible volume and the expansion of gases is rapid. In contrast, in liquids and solids, atoms and molecules are closer together and are quite sensitive to the forces between them.



Atoms and molecules in a gas are typically widely separated, as shown.

Because the forces between them are quite weak at these distances, the properties of a gas depend more on the number of atoms per unit volume and on temperature than on the type of atom.

To get some idea of how pressure, temperature, and volume of a gas are related to one another, consider what happens when you pump air into an initially deflated tire. The tire's volume first increases in direct proportion to the amount of air injected, without much increase in the tire pressure. Once the tire has expanded to nearly its full size, the walls limit volume expansion. If we continue to pump air into it, the pressure increases. The pressure will further increase when the car is driven and the tires move. Most manufacturers specify optimal tire pressure for cold tires. (See [\[link\]](#).)



(a) When air is pumped into a deflated tire, its volume first increases without much increase in pressure. (b) When the tire is filled to a certain point, the tire walls resist further expansion and the pressure increases with

more air. (c) Once the tire is inflated, its pressure increases with temperature.

At room temperatures, collisions between atoms and molecules can be ignored. In this case, the gas is called an ideal gas, in which case the relationship between the pressure, volume, and temperature is given by the equation of state called the ideal gas law.

Note:

Ideal Gas Law

The **ideal gas law** states that

Equation:

$$PV = NkT,$$

where P is the absolute pressure of a gas, V is the volume it occupies, N is the number of atoms and molecules in the gas, and T is its absolute temperature. The constant k is called the **Boltzmann constant** in honor of Austrian physicist Ludwig Boltzmann (1844–1906) and has the value

Equation:

$$k = 1.38 \times 10^{-23} \text{ J/K}.$$

The ideal gas law can be derived from basic principles, but was originally deduced from experimental measurements of Charles' law (that volume occupied by a gas is proportional to temperature at a fixed pressure) and from Boyle's law (that for a fixed temperature, the product PV is a constant). In the ideal gas model, the volume occupied by its atoms and molecules is a negligible fraction of V . The ideal gas law describes the behavior of real gases under most conditions. (Note, for example, that N is the total number of atoms and molecules, independent of the type of gas.)

Let us see how the ideal gas law is consistent with the behavior of filling the tire when it is pumped slowly and the temperature is constant. At first, the pressure P is essentially equal to atmospheric pressure, and the volume V increases in direct proportion to the number of atoms and molecules N put into the tire. Once the volume of the tire is constant, the equation $PV = NkT$ predicts that the pressure should increase in proportion to *the number N of atoms and molecules*.

Example:

Calculating Pressure Changes Due to Temperature Changes: Tire Pressure

Suppose your bicycle tire is fully inflated, with an absolute pressure of 7.00×10^5 Pa (a gauge pressure of just under 90.0 lb/in²) at a temperature of 18.0°C. What is the pressure after its temperature has risen to 35.0°C? Assume that there are no appreciable leaks or changes in volume.

Strategy

The pressure in the tire is changing only because of changes in temperature. First we need to identify what we know and what we want to know, and then identify an equation to solve for the unknown.

We know the initial pressure $P_0 = 7.00 \times 10^5$ Pa, the initial temperature $T_0 = 18.0^\circ\text{C}$, and the final temperature $T_f = 35.0^\circ\text{C}$. We must find the final pressure P_f . How can we use the equation $PV = NkT$? At first, it may seem that not enough information is given, because the volume V and number of atoms N are not specified. What we can do is use the equation twice: $P_0V_0 = NkT_0$ and $P_fV_f = NkT_f$. If we divide P_fV_f by P_0V_0 we can come up with an equation that allows us to solve for P_f .

Equation:

$$\frac{P_f V_f}{P_0 V_0} = \frac{N_f k T_f}{N_0 k T_0}$$

Since the volume is constant, V_f and V_0 are the same and they cancel out. The same is true for N_f and N_0 , and k , which is a constant. Therefore,

Equation:

$$\frac{P_f}{P_0} = \frac{T_f}{T_0}.$$

We can then rearrange this to solve for P_f :

Equation:

$$P_f = P_0 \frac{T_f}{T_0},$$

where the temperature must be in units of kelvins, because T_0 and T_f are absolute temperatures.

Solution

1. Convert temperatures from Celsius to Kelvin.

Equation:

$$\begin{aligned} T_0 &= (18.0 + 273)\text{K} = 291 \text{ K} \\ T_f &= (35.0 + 273)\text{K} = 308 \text{ K} \end{aligned}$$

2. Substitute the known values into the equation.

Equation:

$$P_f = P_0 \frac{T_f}{T_0} = 7.00 \times 10^5 \text{ Pa} \left(\frac{308 \text{ K}}{291 \text{ K}} \right) = 7.41 \times 10^5 \text{ Pa}$$

Discussion

The final temperature is about 6% greater than the original temperature, so the final pressure is about 6% greater as well. Note that *absolute* pressure and *absolute* temperature must be used in the ideal gas law.

Note:

Making Connections: Take-Home Experiment—Refrigerating a Balloon

Inflate a balloon at room temperature. Leave the inflated balloon in the refrigerator overnight. What happens to the balloon, and why?

Example:

Calculating the Number of Molecules in a Cubic Meter of Gas

How many molecules are in a typical object, such as gas in a tire or water in a drink? We can use the ideal gas law to give us an idea of how large N typically is.

Calculate the number of molecules in a cubic meter of gas at standard temperature and pressure (STP), which is defined to be 0°C and atmospheric pressure.

Strategy

Because pressure, volume, and temperature are all specified, we can use the ideal gas law $PV = NkT$, to find N .

Solution

1. Identify the knowns.

Equation:

$$T = 0^\circ\text{C} = 273 \text{ K}$$

$$P = 1.01 \times 10^5 \text{ Pa}$$

$$V = 1.00 \text{ m}^3$$

$$k = 1.38 \times 10^{-23} \text{ J/K}$$

2. Identify the unknown: number of molecules, N .

3. Rearrange the ideal gas law to solve for N .

Equation:

$$PV = NkT$$

$$N = \frac{PV}{kT}$$

4. Substitute the known values into the equation and solve for N .

Equation:

$$N = \frac{PV}{kT} = \frac{(1.01 \times 10^5 \text{ Pa})(1.00 \text{ m}^3)}{(1.38 \times 10^{-23} \text{ J/K})(273 \text{ K})} = 2.68 \times 10^{25} \text{ molecules}$$

Discussion

This number is undeniably large, considering that a gas is mostly empty space. N is huge, even in small volumes. For example, 1 cm^3 of a gas at STP has 2.68×10^{19} molecules in it. Once again, note that N is the same for all types or mixtures of gases.

Moles and Avogadro's Number

It is sometimes convenient to work with a unit other than molecules when measuring the amount of substance. A **mole** (abbreviated mol) is defined to be the amount of a substance that contains as many atoms or molecules as there are atoms in exactly 12 grams (0.012 kg) of carbon-12. The actual number of atoms or molecules in one mole is called **Avogadro's number** (N_A), in recognition of Italian scientist Amedeo Avogadro (1776–1856). He developed the concept of the mole, based on the hypothesis that equal volumes of gas, at the same pressure and temperature, contain equal numbers of molecules. That is, the number is independent of the type of gas. This hypothesis has been confirmed, and the value of Avogadro's number is

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}.$$

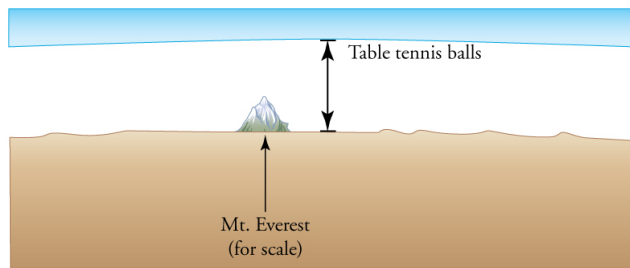
Note:

Avogadro's Number

One mole always contains 6.02×10^{23} particles (atoms or molecules), independent of the element or substance. A mole of any substance has a mass in grams equal to its molecular mass, which can be calculated from the atomic masses given in the periodic table of elements.

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}$$



How big is a mole? On a macroscopic level, one mole of table tennis balls would cover the Earth to a depth of about 40 km.

Exercise:
Check Your Understanding

Problem:

The active ingredient in a Tylenol pill is 325 mg of acetaminophen ($\text{C}_8\text{H}_9\text{NO}_2$). Find the number of active molecules of acetaminophen in a single pill.

Solution:

We first need to calculate the molar mass (the mass of one mole) of acetaminophen. To do this, we need to multiply the number of atoms of each element by the element's atomic mass.

Equation:

$$(8 \text{ moles of carbon})(12 \text{ grams/mole}) + (9 \text{ moles hydrogen})(1 \text{ gram/mole}) \\ + (1 \text{ mole nitrogen})(14 \text{ grams/mole}) + (2 \text{ moles oxygen})(16 \text{ grams/mole}) = 151 \text{ g}$$

Then we need to calculate the number of moles in 325 mg.

Equation:

$$\left(\frac{325 \text{ mg}}{151 \text{ grams/mole}} \right) \left(\frac{1 \text{ gram}}{1000 \text{ mg}} \right) = 2.15 \times 10^{-3} \text{ moles}$$

Then use Avogadro's number to calculate the number of molecules.

Equation:

$$N = (2.15 \times 10^{-3} \text{ moles}) (6.02 \times 10^{23} \text{ molecules/mole}) = 1.30 \times 10^{21} \text{ molecules}$$

Example:

Calculating Moles per Cubic Meter and Liters per Mole

Calculate: (a) the number of moles in 1.00 m^3 of gas at STP, and (b) the number of liters of gas per mole.

Strategy and Solution

(a) We are asked to find the number of moles per cubic meter, and we know from [\[link\]](#) that the number of molecules per cubic meter at STP is 2.68×10^{25} . The number of moles can be found by dividing the number of molecules by Avogadro's number. We let n stand for the number of moles,

Equation:

$$n \text{ mol/m}^3 = \frac{N \text{ molecules/m}^3}{6.02 \times 10^{23} \text{ molecules/mol}} = \frac{2.68 \times 10^{25} \text{ molecules/m}^3}{6.02 \times 10^{23} \text{ molecules/mol}} = 44.5 \text{ mol/m}^3.$$

(b) Using the value obtained for the number of moles in a cubic meter, and converting cubic meters to liters, we obtain

Equation:

$$\frac{(10^3 \text{ L/m}^3)}{44.5 \text{ mol/m}^3} = 22.5 \text{ L/mol.}$$

Discussion

This value is very close to the accepted value of 22.4 L/mol. The slight difference is due to rounding errors caused by using three-digit input. Again this number is the same for all gases. In other words, it is independent of the gas.

The (average) molar weight of air (approximately 80% N₂ and 20% O₂) is $M = 28.8 \text{ g}$. Thus the mass of one cubic meter of air is 1.28 kg. If a living room has dimensions $5 \text{ m} \times 5 \text{ m} \times 3 \text{ m}$, the mass of air inside the room is 96 kg, which is the typical mass of a human.

Exercise:

Check Your Understanding

Problem:

The density of air at standard conditions ($P = 1 \text{ atm}$ and $T = 20^\circ\text{C}$) is 1.28 kg/m^3 . At what pressure is the density 0.64 kg/m^3 if the temperature and number of molecules are kept constant?

Solution:

The best way to approach this question is to think about what is happening. If the density drops to half its original value and no molecules are lost, then the volume must double. If we look at the equation $PV = NkT$, we see that when the temperature is constant, the pressure is inversely proportional to volume. Therefore, if the volume doubles, the pressure must drop to half its original value, and $P_f = 0.50 \text{ atm}$.

The Ideal Gas Law Restated Using Moles

A very common expression of the ideal gas law uses the number of moles, n , rather than the number of atoms and molecules, N . We start from the ideal gas law,

Equation:

$$PV = NkT,$$

and multiply and divide the equation by Avogadro's number N_A . This gives

Equation:

$$PV = \frac{N}{N_A} N_A k T.$$

Note that $n = N/N_A$ is the number of moles. We define the universal gas constant $R = N_A k$, and obtain the ideal gas law in terms of moles.

Note:

Ideal Gas Law (in terms of moles)

The ideal gas law (in terms of moles) is

Equation:

$$PV = nRT.$$

The numerical value of R in SI units is

Equation:

$$R = N_A k = (6.02 \times 10^{23} \text{ mol}^{-1})(1.38 \times 10^{-23} \text{ J/K}) = 8.31 \text{ J/mol} \cdot \text{K}.$$

In other units,

Equation:

$$R = 1.99 \text{ cal/mol} \cdot \text{K}$$

$$R = 0.0821 \text{ L} \cdot \text{atm/mol} \cdot \text{K}.$$

You can use whichever value of R is most convenient for a particular problem.

Example:

Calculating Number of Moles: Gas in a Bike Tire

How many moles of gas are in a bike tire with a volume of $2.00 \times 10^{-3} \text{ m}^3$ (2.00 L), a pressure of $7.00 \times 10^5 \text{ Pa}$ (a gauge pressure of just under 90.0 lb/in²), and at a temperature of 18.0°C?

Strategy

Identify the knowns and unknowns, and choose an equation to solve for the unknown. In this case, we solve the ideal gas law, $PV = nRT$, for the number of moles n .

Solution

1. Identify the knowns.

Equation:

$$P = 7.00 \times 10^5 \text{ Pa}$$

$$V = 2.00 \times 10^{-3} \text{ m}^3$$

$$T = 18.0^\circ\text{C} = 291 \text{ K}$$

$$R = 8.31 \text{ J/mol} \cdot \text{K}$$

2. Rearrange the equation to solve for n and substitute known values.

Equation:

$$\begin{aligned} n &= \frac{PV}{RT} = \frac{(7.00 \times 10^5 \text{ Pa})(2.00 \times 10^{-3} \text{ m}^3)}{(8.31 \text{ J/mol}\cdot\text{K})(291 \text{ K})} \\ &= 0.579 \text{ mol} \end{aligned}$$

Discussion

The most convenient choice for R in this case is $8.31 \text{ J/mol}\cdot\text{K}$, because our known quantities are in SI units. The pressure and temperature are obtained from the initial conditions in [\[link\]](#), but we would get the same answer if we used the final values.

The ideal gas law can be considered to be another manifestation of the law of conservation of energy (see [Conservation of Energy](#)). Work done on a gas results in an increase in its energy, increasing pressure and/or temperature, or decreasing volume. This increased energy can also be viewed as increased internal kinetic energy, given the gas's atoms and molecules.

The Ideal Gas Law and Energy

Let us now examine the role of energy in the behavior of gases. When you inflate a bike tire by hand, you do work by repeatedly exerting a force through a distance. This energy goes into increasing the pressure of air inside the tire and increasing the temperature of the pump and the air.

The ideal gas law is closely related to energy: the units on both sides are joules. The right-hand side of the ideal gas law in $PV = NkT$ is NkT . This term is roughly the amount of translational kinetic energy of N atoms or molecules at an absolute temperature T , as we shall see formally in [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#). The left-hand side of the ideal gas law is PV , which also has the units of joules. We know from our study of fluids that pressure is one type of potential energy per unit volume, so pressure multiplied by volume is energy. The important point is that there is energy in a gas related to both its pressure and its volume. The energy can be changed when the gas is doing work as it expands—something we explore in [Heat and Heat Transfer Methods](#)—similar to what occurs in gasoline or steam engines and turbines.

Note:

Problem-Solving Strategy: The Ideal Gas Law

Step 1 Examine the situation to determine that an ideal gas is involved. Most gases are nearly ideal.

Step 2 Make a list of what quantities are given, or can be inferred from the problem as stated (identify the known quantities). Convert known values into proper SI units (K for temperature, Pa for pressure, m^3 for volume, molecules for N , and moles for n).

Step 3 Identify exactly what needs to be determined in the problem (identify the unknown quantities). A written list is useful.

Step 4 Determine whether the number of molecules or the number of moles is known, in order to decide which form of the ideal gas law to use. The first form is $PV = NkT$ and involves N , the number of atoms or molecules. The second form is $PV = nRT$ and involves n , the number of moles.

Step 5 Solve the ideal gas law for the quantity to be determined (the unknown quantity). You may need to take a ratio of final states to initial states to eliminate the unknown quantities that are kept fixed.

Step 6 Substitute the known quantities, along with their units, into the appropriate equation, and obtain numerical solutions complete with units. Be certain to use absolute temperature and absolute pressure.

Step 7 Check the answer to see if it is reasonable: Does it make sense?

Exercise:

Check Your Understanding

Problem:

Liquids and solids have densities about 1000 times greater than gases. Explain how this implies that the distances between atoms and molecules in gases are about 10 times greater than the size of their atoms and molecules.

Solution:

Atoms and molecules are close together in solids and liquids. In gases they are separated by empty space. Thus gases have lower densities than liquids and solids. Density is mass per unit volume, and volume is related to the size of a body (such as a sphere) cubed. So if the distance between atoms and molecules increases by a factor of 10, then the volume occupied increases by a factor of 1000, and the density decreases by a factor of 1000.

Section Summary

- The ideal gas law relates the pressure and volume of a gas to the number of gas molecules and the temperature of the gas.
- The ideal gas law can be written in terms of the number of molecules of gas:

Equation:

$$PV = NkT,$$

where P is pressure, V is volume, T is temperature, N is number of molecules, and k is the Boltzmann constant

Equation:

$$k = 1.38 \times 10^{-23} \text{ J/K}.$$

- A mole is the number of atoms in a 12-g sample of carbon-12.
- The number of molecules in a mole is called Avogadro's number N_A ,

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}.$$

- A mole of any substance has a mass in grams equal to its molecular weight, which can be determined from the periodic table of elements.
- The ideal gas law can also be written and solved in terms of the number of moles of gas:

Equation:

$$PV = nRT,$$

where n is number of moles and R is the universal gas constant,

Equation:

$$R = 8.31 \text{ J/mol} \cdot \text{K}.$$

- The ideal gas law is generally valid at temperatures well above the boiling temperature.

Conceptual Questions

Exercise:

Problem:

Find out the human population of Earth. Is there a mole of people inhabiting Earth? If the average mass of a person is 60 kg, calculate the mass of a mole of people. How does the mass of a mole of people compare with the mass of Earth?

Exercise:

Problem:

Under what circumstances would you expect a gas to behave significantly differently than predicted by the ideal gas law?

Exercise:

Problem:

A constant-volume gas thermometer contains a fixed amount of gas. What property of the gas is measured to indicate its temperature?

Problems & Exercises

Exercise:

Problem:

The gauge pressure in your car tires is $2.50 \times 10^5 \text{ N/m}^2$ at a temperature of 35.0°C when you drive it onto a ferry boat to Alaska. What is their gauge pressure later, when their temperature has dropped to -40.0°C ?

Solution:

1.62 atm

Exercise:**Problem:**

Convert an absolute pressure of $7.00 \times 10^5 \text{ N/m}^2$ to gauge pressure in lb/in^2 . (This value was stated to be just less than 90.0 lb/in^2 in [\[link\]](#). Is it?)

Exercise:**Problem:**

Suppose a gas-filled incandescent light bulb is manufactured so that the gas inside the bulb is at atmospheric pressure when the bulb has a temperature of 20.0°C . (a) Find the gauge pressure inside such a bulb when it is hot, assuming its average temperature is 60.0°C (an approximation) and neglecting any change in volume due to thermal expansion or gas leaks. (b) The actual final pressure for the light bulb will be less than calculated in part (a) because the glass bulb will expand. What will the actual final pressure be, taking this into account? Is this a negligible difference?

Solution:

(a) 0.136 atm

(b) 0.135 atm. The difference between this value and the value from part (a) is negligible.

Exercise:**Problem:**

Large helium-filled balloons are used to lift scientific equipment to high altitudes. (a) What is the pressure inside such a balloon if it starts out at sea level with a temperature of 10.0°C and rises to an altitude where its volume is twenty times the original volume and its temperature is -50.0°C ? (b) What is the gauge pressure? (Assume atmospheric pressure is constant.)

Exercise:**Problem:**

Confirm that the units of nRT are those of energy for each value of R : (a) $8.31 \text{ J/mol} \cdot \text{K}$, (b) $1.99 \text{ cal/mol} \cdot \text{K}$, and (c) $0.0821 \text{ L} \cdot \text{atm/mol} \cdot \text{K}$.

Solution:

(a) $nRT = (\text{mol})(\text{J/mol} \cdot \text{K})(\text{K}) = \text{J}$

(b) $nRT = (\text{mol})(\text{cal/mol} \cdot \text{K})(\text{K}) = \text{cal}$

$$\begin{aligned}
 nRT &= (\text{mol})(\text{L} \cdot \text{atm}/\text{mol} \cdot \text{K})(\text{K}) \\
 \text{(c)} \quad &= \text{L} \cdot \text{atm} = (\text{m}^3)(\text{N}/\text{m}^2) \\
 &= \text{N} \cdot \text{m} = \text{J}
 \end{aligned}$$

Exercise:

Problem:

In the text, it was shown that $N/V = 2.68 \times 10^{25} \text{ m}^{-3}$ for gas at STP. (a) Show that this quantity is equivalent to $N/V = 2.68 \times 10^{19} \text{ cm}^{-3}$, as stated. (b) About how many atoms are there in one μm^3 (a cubic micrometer) at STP? (c) What does your answer to part (b) imply about the separation of atoms and molecules?

Exercise:

Problem:

Calculate the number of moles in the 2.00-L volume of air in the lungs of the average person. Note that the air is at 37.0°C (body temperature).

Solution:

$$7.86 \times 10^{-2} \text{ mol}$$

Exercise:

Problem:

An airplane passenger has 100 cm^3 of air in his stomach just before the plane takes off from a sea-level airport. What volume will the air have at cruising altitude if cabin pressure drops to $7.50 \times 10^4 \text{ N}/\text{m}^2$?

Exercise:

Problem:

(a) What is the volume (in km^3) of Avogadro's number of sand grains if each grain is a cube and has sides that are 1.0 mm long? (b) How many kilometers of beaches in length would this cover if the beach averages 100 m in width and 10.0 m in depth? Neglect air spaces between grains.

Solution:

$$\text{(a)} \quad 6.02 \times 10^5 \text{ km}^3$$

$$\text{(b)} \quad 6.02 \times 10^8 \text{ km}$$

Exercise:

Problem:

An expensive vacuum system can achieve a pressure as low as $1.00 \times 10^{-7} \text{ N/m}^2$ at 20°C . How many atoms are there in a cubic centimeter at this pressure and temperature?

Exercise:**Problem:**

The number density of gas atoms at a certain location in the space above our planet is about $1.00 \times 10^{11} \text{ m}^{-3}$, and the pressure is $2.75 \times 10^{-10} \text{ N/m}^2$ in this space. What is the temperature there?

Solution:

-73.9°C

Exercise:**Problem:**

A bicycle tire has a pressure of $7.00 \times 10^5 \text{ N/m}^2$ at a temperature of 18.0°C and contains 2.00 L of gas. What will its pressure be if you let out an amount of air that has a volume of 100 cm^3 at atmospheric pressure? Assume tire temperature and volume remain constant.

Exercise:**Problem:**

A high-pressure gas cylinder contains 50.0 L of toxic gas at a pressure of $1.40 \times 10^7 \text{ N/m}^2$ and a temperature of 25.0°C . Its valve leaks after the cylinder is dropped. The cylinder is cooled to dry ice temperature (-78.5°C) to reduce the leak rate and pressure so that it can be safely repaired. (a) What is the final pressure in the tank, assuming a negligible amount of gas leaks while being cooled and that there is no phase change? (b) What is the final pressure if one-tenth of the gas escapes? (c) To what temperature must the tank be cooled to reduce the pressure to 1.00 atm (assuming the gas does not change phase and that there is no leakage during cooling)? (d) Does cooling the tank appear to be a practical solution?

Solution:

(a) $9.14 \times 10^6 \text{ N/m}^2$

(b) $8.23 \times 10^6 \text{ N/m}^2$

(c) 2.16 K

(d) No. The final temperature needed is much too low to be easily achieved for a large object.

Exercise:

Problem:

Find the number of moles in 2.00 L of gas at 35.0°C and under $7.41 \times 10^7 \text{ N/m}^2$ of pressure.

Exercise:**Problem:**

Calculate the depth to which Avogadro's number of table tennis balls would cover Earth. Each ball has a diameter of 3.75 cm. Assume the space between balls adds an extra 25.0% to their volume and assume they are not crushed by their own weight.

Solution:

41 km

Exercise:**Problem:**

(a) What is the gauge pressure in a 25.0°C car tire containing 3.60 mol of gas in a 30.0 L volume? (b) What will its gauge pressure be if you add 1.00 L of gas originally at atmospheric pressure and 25.0°C? Assume the temperature returns to 25.0°C and the volume remains constant.

Exercise:**Problem:**

(a) In the deep space between galaxies, the density of atoms is as low as 10^6 atoms/m^3 , and the temperature is a frigid 2.7 K. What is the pressure? (b) What volume (in m^3) is occupied by 1 mol of gas? (c) If this volume is a cube, what is the length of its sides in kilometers?

Solution:

(a) $3.7 \times 10^{-17} \text{ Pa}$

(b) $6.0 \times 10^{17} \text{ m}^3$

(c) $8.4 \times 10^2 \text{ km}$

Glossary

ideal gas law

the physical law that relates the pressure and volume of a gas to the number of gas molecules or number of moles of gas and the temperature of the gas

Boltzmann constant

k , a physical constant that relates energy to temperature; $k = 1.38 \times 10^{-23} \text{ J/K}$

Avogadro's number

N_A , the number of molecules or atoms in one mole of a substance; $N_A = 6.02 \times 10^{23}$
particles/mole

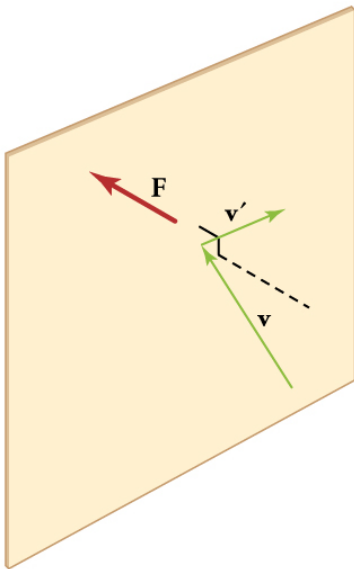
mole

the quantity of a substance whose mass (in grams) is equal to its molecular mass

Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature

- Express the ideal gas law in terms of molecular mass and velocity.
- Define thermal energy.
- Calculate the kinetic energy of a gas molecule, given its temperature.
- Describe the relationship between the temperature of a gas and the kinetic energy of atoms and molecules.
- Describe the distribution of speeds of molecules in a gas.

We have developed macroscopic definitions of pressure and temperature. Pressure is the force divided by the area on which the force is exerted, and temperature is measured with a thermometer. We gain a better understanding of pressure and temperature from the kinetic theory of gases, which assumes that atoms and molecules are in continuous random motion.



When a molecule collides with a rigid wall, the component of its momentum perpendicular to the wall is reversed. A force is thus exerted on the wall, creating pressure.

[\[link\]](#) shows an elastic collision of a gas molecule with the wall of a container, so that it exerts a force on the wall (by Newton's third law). Because a huge number of molecules will collide with the wall in a short time, we observe an average force per unit area. These collisions are the source of pressure in a gas. As the number of molecules increases, the number of collisions and thus the pressure increase. Similarly, the gas pressure is higher if the average velocity of molecules is higher. The actual relationship is derived in the [Things Great and Small](#) feature below. The following relationship is found:

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2},$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.

What can we learn from this atomic and molecular version of the ideal gas law? We can derive a relationship between temperature and the average translational kinetic energy of molecules in a gas. Recall the previous expression of the ideal gas law:

Equation:

$$PV = NkT.$$

Equating the right-hand side of this equation with the right-hand side of $PV = \frac{1}{3}Nm\overline{v^2}$ gives

Equation:

$$\frac{1}{3}Nm\overline{v^2} = NkT.$$

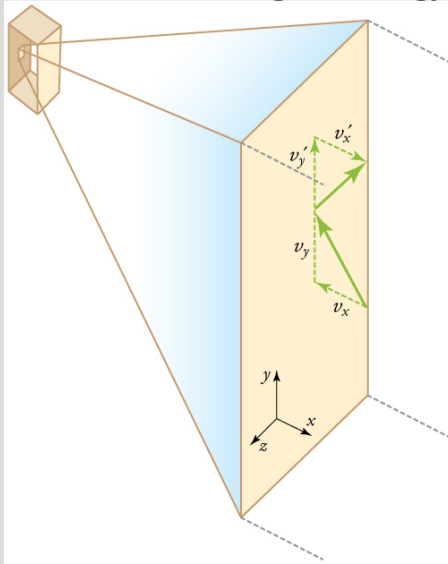
Note:

Making Connections: Things Great and Small—Atomic and Molecular Origin of Pressure in a Gas

[\[link\]](#) shows a box filled with a gas. We know from our previous discussions that putting more gas into the box produces greater pressure, and that increasing the temperature of the gas also produces a greater pressure. But why should increasing the temperature of the gas increase the pressure in the box? A look at the atomic and

molecular scale gives us some answers, and an alternative expression for the ideal gas law.

The figure shows an expanded view of an elastic collision of a gas molecule with the wall of a container. Calculating the average force exerted by such molecules will lead us to the ideal gas law, and to the connection between temperature and molecular kinetic energy. We assume that a molecule is small compared with the separation of molecules in the gas, and that its interaction with other molecules can be ignored. We also assume the wall is rigid and that the molecule's direction changes, but that its speed remains constant (and hence its kinetic energy and the magnitude of its momentum remain constant as well). This assumption is not always valid, but the same result is obtained with a more detailed description of the molecule's exchange of energy and momentum with the wall.



Gas in a box exerts an outward pressure on its walls. A molecule colliding with a rigid wall has the direction of its velocity and momentum in the x -direction reversed. This direction is perpendicular to the wall.

The components of its velocity momentum in the y - and z -directions are not changed, which means there is no force parallel to the wall.

If the molecule's velocity changes in the x -direction, its momentum changes from $-mv_x$ to $+mv_x$. Thus, its change in momentum is $\Delta mv = +mv_x - (-mv_x) = 2mv_x$. The force exerted on the molecule is given by

Equation:

$$F = \frac{\Delta p}{\Delta t} = \frac{2mv_x}{\Delta t}.$$

There is no force between the wall and the molecule until the molecule hits the wall. During the short time of the collision, the force between the molecule and wall is relatively large. We are looking for an average force; we take Δt to be the average time between collisions of the molecule with this wall. It is the time it would take the molecule to go across the box and back (a distance $2l$) at a speed of v_x . Thus $\Delta t = 2l/v_x$, and the expression for the force becomes

Equation:

$$F = \frac{2mv_x}{2l/v_x} = \frac{mv_x^2}{l}.$$

This force is due to *one* molecule. We multiply by the number of molecules N and use their average squared velocity to find the force

Equation:

$$F = N \frac{\overline{mv_x^2}}{l},$$

where the bar over a quantity means its average value. We would like to have the force in terms of the speed v , rather than the x -component of the velocity. We note that the total velocity squared is the sum of the squares of its components, so that

Equation:

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}.$$

Because the velocities are random, their average components in all directions are the same:

Equation:

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}.$$

Thus,

Equation:

$$\overline{v^2} = 3\overline{v_x^2},$$

or

Equation:

$$\overline{v_x^2} = \frac{1}{3}\overline{v^2}.$$

Substituting $\frac{1}{3}\overline{v^2}$ into the expression for F gives

Equation:

$$F = N \frac{m\overline{v^2}}{3l}.$$

The pressure is F/A , so that we obtain

Equation:

$$P = \frac{F}{A} = N \frac{m\overline{v^2}}{3Al} = \frac{1}{3} \frac{Nm\overline{v^2}}{V},$$

where we used $V = Al$ for the volume. This gives the important result.

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2}$$

This equation is another expression of the ideal gas law.

We can get the average kinetic energy of a molecule, $\frac{1}{2}mv^2$, from the right-hand side of the equation by canceling N and multiplying by $3/2$. This calculation produces the result that the average kinetic energy of a molecule is directly related to absolute temperature.

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

The average translational kinetic energy of a molecule, \overline{KE} , is called **thermal energy**. The equation $\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$ is a molecular interpretation of temperature, and it has been found to be valid for gases and reasonably accurate in liquids and solids. It is another definition of temperature based on an expression of the molecular energy.

It is sometimes useful to rearrange $\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$, and solve for the average speed of molecules in a gas in terms of temperature,

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}},$$

where v_{rms} stands for root-mean-square (rms) speed.

Example:

Calculating Kinetic Energy and Speed of a Gas Molecule

(a) What is the average kinetic energy of a gas molecule at 20.0°C (room temperature)? (b) Find the rms speed of a nitrogen molecule (N_2) at this temperature.

Strategy for (a)

The known in the equation for the average kinetic energy is the temperature.

Equation:

$$\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

Before substituting values into this equation, we must convert the given temperature to kelvins. This conversion gives $T = (20.0 + 273) \text{ K} = 293 \text{ K}$.

Solution for (a)

The temperature alone is sufficient to find the average translational kinetic energy. Substituting the temperature into the translational kinetic energy equation gives

Equation:

$$\overline{KE} = \frac{3}{2}kT = \frac{3}{2}(1.38 \times 10^{-23} \text{ J/K})(293 \text{ K}) = 6.07 \times 10^{-21} \text{ J}.$$

Strategy for (b)

Finding the rms speed of a nitrogen molecule involves a straightforward calculation using the equation

Equation:

$$\sqrt{v^2} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}},$$

but we must first find the mass of a nitrogen molecule. Using the molecular mass of nitrogen N_2 from the periodic table,

Equation:

$$m = \frac{2(14.0067) \times 10^{-3} \text{ kg/mol}}{6.02 \times 10^{23} \text{ mol}^{-1}} = 4.65 \times 10^{-26} \text{ kg}.$$

Solution for (b)

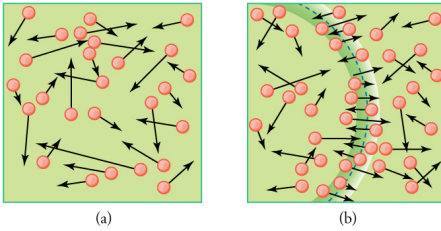
Substituting this mass and the value for k into the equation for v_{rms} yields

Equation:

$$v_{\text{rms}} = \sqrt{\frac{3kT}{m}} = \sqrt{\frac{3(1.38 \times 10^{-23} \text{ J/K})(293 \text{ K})}{4.65 \times 10^{-26} \text{ kg}}} = 511 \text{ m/s}.$$

Discussion

Note that the average kinetic energy of the molecule is independent of the type of molecule. The average translational kinetic energy depends only on absolute temperature. The kinetic energy is very small compared to macroscopic energies, so that we do not feel when an air molecule is hitting our skin. The rms velocity of the nitrogen molecule is surprisingly large. These large molecular velocities do not yield macroscopic movement of air, since the molecules move in all directions with equal likelihood. The *mean free path* (the distance a molecule can move on average between collisions) of molecules in air is very small, and so the molecules move rapidly but do not get very far in a second. The high value for rms speed is reflected in the speed of sound, however, which is about 340 m/s at room temperature. The faster the rms speed of air molecules, the faster that sound vibrations can be transferred through the air. The speed of sound increases with temperature and is greater in gases with small molecular masses, such as helium. (See [\[link\]](#).)



(a) There are many molecules moving so fast in an ordinary gas that they collide a billion times every second. (b) Individual molecules do not move very far in a small amount of time, but disturbances like sound waves are transmitted at speeds related to the molecular speeds.

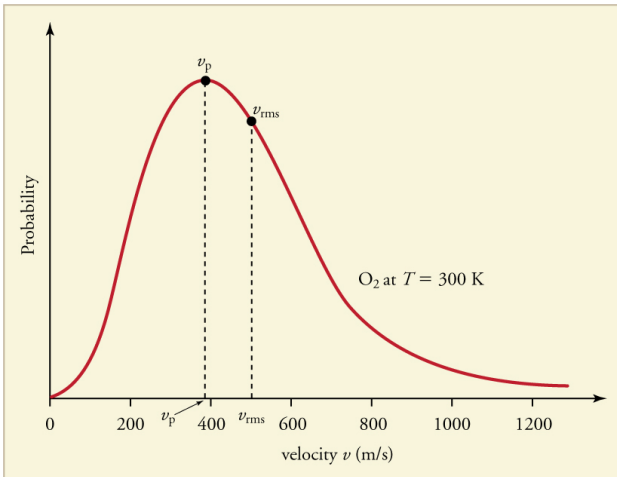
Note:

Making Connections: Historical Note—Kinetic Theory of Gases

The kinetic theory of gases was developed by Daniel Bernoulli (1700–1782), who is best known in physics for his work on fluid flow (hydrodynamics). Bernoulli's work predates the atomistic view of matter established by Dalton.

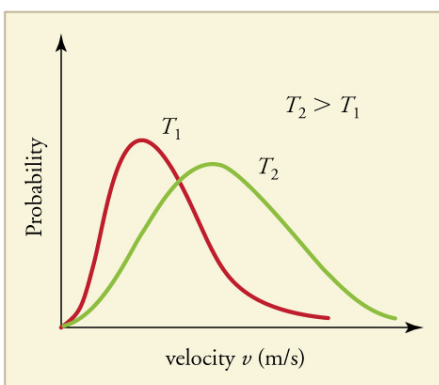
Distribution of Molecular Speeds

The motion of molecules in a gas is random in magnitude and direction for individual molecules, but a gas of many molecules has a predictable distribution of molecular speeds. This distribution is called the *Maxwell-Boltzmann distribution*, after its originators, who calculated it based on kinetic theory, and has since been confirmed experimentally. (See [\[link\]](#).) The distribution has a long tail, because a few molecules may go several times the rms speed. The most probable speed v_p is less than the rms speed v_{rms} . [\[link\]](#) shows that the curve is shifted to higher speeds at higher temperatures, with a broader range of speeds.



The Maxwell-Boltzmann distribution of molecular speeds in an ideal gas. The most likely speed v_p is less than the rms speed v_{rms} . Although very high speeds are possible, only a tiny fraction of the molecules have speeds that are an order of magnitude greater than v_{rms} .

The distribution of thermal speeds depends strongly on temperature. As temperature increases, the speeds are shifted to higher values and the distribution is broadened.



The Maxwell-Boltzmann distribution is shifted to

higher speeds and is broadened at higher temperatures.

What is the implication of the change in distribution with temperature shown in [\[link\]](#) for humans? All other things being equal, if a person has a fever, he or she is likely to lose more water molecules, particularly from linings along moist cavities such as the lungs and mouth, creating a dry sensation in the mouth.

Example:

Calculating Temperature: Escape Velocity of Helium Atoms

In order to escape Earth's gravity, an object near the top of the atmosphere (at an altitude of 100 km) must travel away from Earth at 11.1 km/s. This speed is called the *escape velocity*. At what temperature would helium atoms have an rms speed equal to the escape velocity?

Strategy

Identify the knowns and unknowns and determine which equations to use to solve the problem.

Solution

1. Identify the knowns: v is the escape velocity, 11.1 km/s.
2. Identify the unknowns: We need to solve for temperature, T . We also need to solve for the mass m of the helium atom.
3. Determine which equations are needed.

- To solve for mass m of the helium atom, we can use information from the periodic table:

Equation:

$$m = \frac{\text{molar mass}}{\text{number of atoms per mole}}.$$

- To solve for temperature T , we can rearrange either

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

or

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}}$$

to yield

Equation:

$$T = \frac{m\overline{v^2}}{3k},$$

where k is the Boltzmann constant and m is the mass of a helium atom.

4. Plug the known values into the equations and solve for the unknowns.

Equation:

$$m = \frac{\text{molar mass}}{\text{number of atoms per mole}} = \frac{4.0026 \times 10^{-3} \text{ kg/mol}}{6.02 \times 10^{23} \text{ mol}} = 6.65 \times 10^{-27} \text{ kg}$$

Equation:

$$T = \frac{(6.65 \times 10^{-27} \text{ kg})(11.1 \times 10^3 \text{ m/s})^2}{3(1.38 \times 10^{-23} \text{ J/K})} = 1.98 \times 10^4 \text{ K}$$

Discussion

This temperature is much higher than atmospheric temperature, which is approximately 250 K (-25°C or -10°F) at high altitude. Very few helium atoms are left in the atmosphere, but there were many when the atmosphere was formed. The reason for the loss of helium atoms is that there are a small number of helium atoms with speeds higher than Earth's escape velocity even at normal temperatures. The speed of a helium atom changes from one instant to the next, so that at any instant, there is a small, but nonzero chance that the speed is greater than the escape speed and the molecule escapes from Earth's gravitational pull. Heavier molecules, such as oxygen, nitrogen, and water (very little of which reach a very high altitude), have smaller rms speeds, and so it is much less likely that any of them will have speeds greater than the escape velocity. In fact, so few have speeds above the escape velocity that billions of years are required to lose significant amounts of the atmosphere. [\[link\]](#) shows the impact of a lack of an atmosphere on the Moon. Because the gravitational pull of the Moon is much weaker, it has lost almost its

entire atmosphere. The comparison between Earth and the Moon is discussed in this chapter's Problems and Exercises.



This photograph of Apollo 17 Commander Eugene Cernan driving the lunar rover on the Moon in 1972 looks as though it was taken at night with a large spotlight. In fact, the light is coming from the Sun. Because the acceleration due to gravity on the Moon is so low (about $\frac{1}{6}$ that of Earth), the Moon's escape velocity is much smaller. As a result, gas molecules escape very easily from the Moon, leaving it with virtually no atmosphere. Even during the daytime, the sky is black because there is no gas to scatter sunlight. (credit: Harrison H. Schmitt/NASA)

Exercise:**Check Your Understanding****Problem:**

If you consider a very small object such as a grain of pollen, in a gas, then the number of atoms and molecules striking its surface would also be relatively small. Would the grain of pollen experience any fluctuations in pressure due to statistical fluctuations in the number of gas atoms and molecules striking it in a given amount of time?

Solution:

Yes. Such fluctuations actually occur for a body of any size in a gas, but since the numbers of atoms and molecules are immense for macroscopic bodies, the fluctuations are a tiny percentage of the number of collisions, and the averages spoken of in this section vary imperceptibly. Roughly speaking the fluctuations are proportional to the inverse square root of the number of collisions, so for small bodies they can become significant. This was actually observed in the 19th century for pollen grains in water, and is known as the Brownian effect.

Note:

PhET Explorations: Gas Properties

Pump gas molecules into a box and see what happens as you change the volume, add or remove heat, change gravity, and more. Measure the temperature and pressure, and discover how the properties of the gas vary in relation to each other.

[Gas](#)
[Propertie](#)
[s](#)

Section Summary

- Kinetic theory is the atomistic description of gases as well as liquids and solids.

- Kinetic theory models the properties of matter in terms of continuous random motion of atoms and molecules.
- The ideal gas law can also be expressed as

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2},$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.

- Thermal energy is defined to be the average translational kinetic energy $\overline{\text{KE}}$ of an atom or molecule.
- The temperature of gases is proportional to the average translational kinetic energy of atoms and molecules.

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

or

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}}.$$

- The motion of individual molecules in a gas is random in magnitude and direction. However, a gas of many molecules has a predictable distribution of molecular speeds, known as the *Maxwell-Boltzmann distribution*.

Conceptual Questions

Exercise:

Problem:

How is momentum related to the pressure exerted by a gas? Explain on the atomic and molecular level, considering the behavior of atoms and molecules.

Problems & Exercises

Exercise:**Problem:**

Some incandescent light bulbs are filled with argon gas. What is v_{rms} for argon atoms near the filament, assuming their temperature is 2500 K?

Solution:

$$1.25 \times 10^3 \text{ m/s}$$

Exercise:**Problem:**

Average atomic and molecular speeds (v_{rms}) are large, even at low temperatures. What is v_{rms} for helium atoms at 5.00 K, just one degree above helium's liquefaction temperature?

Exercise:**Problem:**

(a) What is the average kinetic energy in joules of hydrogen atoms on the 5500°C surface of the Sun? (b) What is the average kinetic energy of helium atoms in a region of the solar corona where the temperature is $6.00 \times 10^5 \text{ K}$?

Solution:

(a) $1.20 \times 10^{-19} \text{ J}$

(b) $1.24 \times 10^{-17} \text{ J}$

Exercise:**Problem:**

The escape velocity of any object from Earth is 11.2 km/s. (a) Express this speed in m/s and km/h. (b) At what temperature would oxygen molecules (molecular mass is equal to 32.0 g/mol) have an average velocity v_{rms} equal to Earth's escape velocity of 11.1 km/s?

Exercise:

Problem:

The escape velocity from the Moon is much smaller than from Earth and is only 2.38 km/s. At what temperature would hydrogen molecules (molecular mass is equal to 2.016 g/mol) have an average velocity v_{rms} equal to the Moon's escape velocity?

Solution:

458 K

Exercise:**Problem:**

Nuclear fusion, the energy source of the Sun, hydrogen bombs, and fusion reactors, occurs much more readily when the average kinetic energy of the atoms is high—that is, at high temperatures. Suppose you want the atoms in your fusion experiment to have average kinetic energies of 6.40×10^{-14} J. What temperature is needed?

Exercise:**Problem:**

Suppose that the average velocity (v_{rms}) of carbon dioxide molecules (molecular mass is equal to 44.0 g/mol) in a flame is found to be 1.05×10^5 m/s. What temperature does this represent?

Solution:

1.95×10^7 K

Exercise:**Problem:**

Hydrogen molecules (molecular mass is equal to 2.016 g/mol) have an average velocity v_{rms} equal to 193 m/s. What is the temperature?

Exercise:

Problem:

Much of the gas near the Sun is atomic hydrogen. Its temperature would have to be 1.5×10^7 K for the average velocity v_{rms} to equal the escape velocity from the Sun. What is that velocity?

Solution:

$$6.09 \times 10^5 \text{ m/s}$$

Exercise:**Problem:**

There are two important isotopes of uranium— ^{235}U and ^{238}U ; these isotopes are nearly identical chemically but have different atomic masses. Only ^{235}U is very useful in nuclear reactors. One of the techniques for separating them (gas diffusion) is based on the different average velocities v_{rms} of uranium hexafluoride gas, UF_6 . (a) The molecular masses for $^{235}\text{U UF}_6$ and $^{238}\text{U UF}_6$ are 349.0 g/mol and 352.0 g/mol, respectively. What is the ratio of their average velocities? (b) At what temperature would their average velocities differ by 1.00 m/s? (c) Do your answers in this problem imply that this technique may be difficult?

Glossary

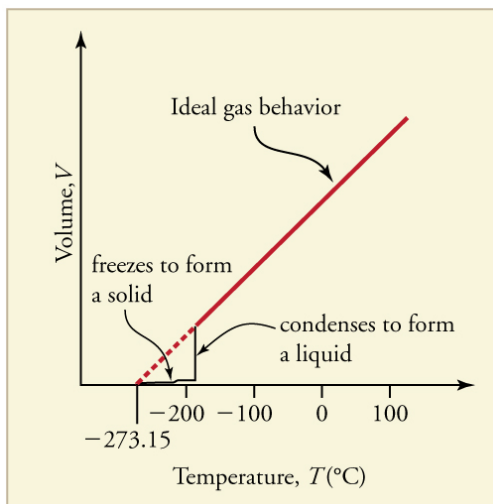
thermal energy

$\overline{\text{KE}}$, the average translational kinetic energy of a molecule

Phase Changes

- Interpret a phase diagram.
- State Dalton's law.
- Identify and describe the triple point of a gas from its phase diagram.
- Describe the state of equilibrium between a liquid and a gas, a liquid and a solid, and a gas and a solid.

Up to now, we have considered the behavior of ideal gases. Real gases are like ideal gases at high temperatures. At lower temperatures, however, the interactions between the molecules and their volumes cannot be ignored. The molecules are very close (condensation occurs) and there is a dramatic decrease in volume, as seen in [\[link\]](#). The substance changes from a gas to a liquid. When a liquid is cooled to even lower temperatures, it becomes a solid. The volume never reaches zero because of the finite volume of the molecules.



A sketch of volume versus temperature for a real gas at constant pressure. The linear (straight line) part of the graph represents ideal gas behavior—volume and temperature are directly and positively related and

the line extrapolates to zero volume at -273.15°C , or absolute zero. When the gas becomes a liquid, however, the volume actually decreases precipitously at the liquefaction point. The volume decreases slightly once the substance is solid, but it never becomes zero.

High pressure may also cause a gas to change phase to a liquid. Carbon dioxide, for example, is a gas at room temperature and atmospheric pressure, but becomes a liquid under sufficiently high pressure. If the pressure is reduced, the temperature drops and the liquid carbon dioxide solidifies into a snow-like substance at the temperature -78°C . Solid CO_2 is called “dry ice.” Another example of a gas that can be in a liquid phase is liquid nitrogen (LN_2). LN_2 is made by liquefaction of atmospheric air (through compression and cooling). It boils at 77 K (-196°C) at atmospheric pressure. LN_2 is useful as a refrigerant and allows for the preservation of blood, sperm, and other biological materials. It is also used to reduce noise in electronic sensors and equipment, and to help cool down their current-carrying wires. In dermatology, LN_2 is used to freeze and painlessly remove warts and other growths from the skin.

PV Diagrams

We can examine aspects of the behavior of a substance by plotting a graph of pressure versus volume, called a **PV diagram**. When the substance behaves like an ideal gas, the ideal gas law describes the relationship between its pressure and volume. That is,

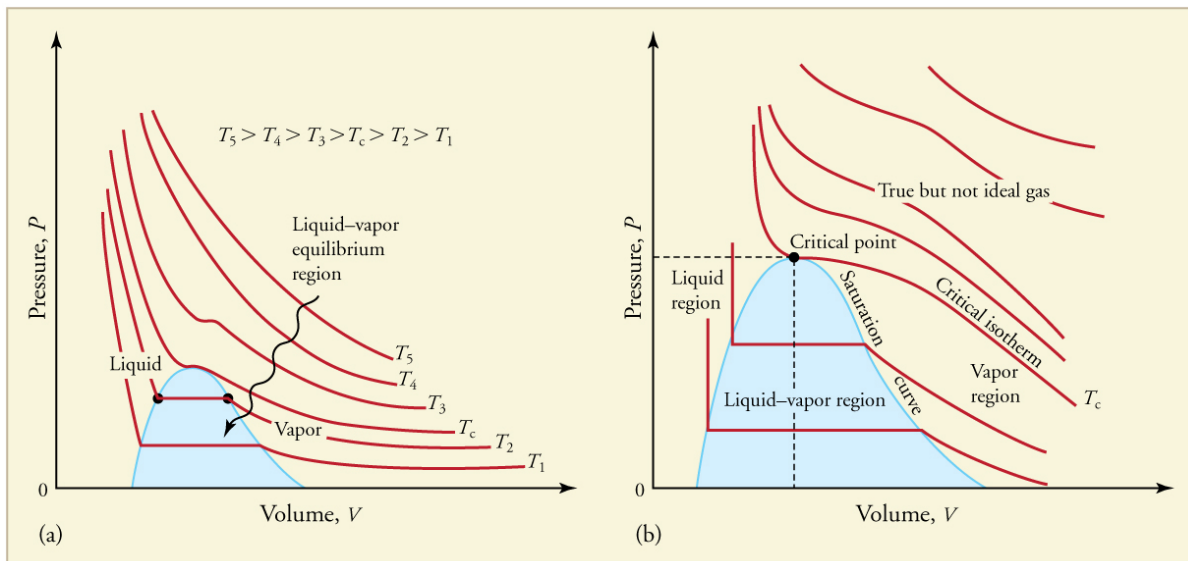
Equation:

$$PV = NkT \text{ (ideal gas).}$$

Now, assuming the number of molecules and the temperature are fixed,
Equation:

$$PV = \text{constant (ideal gas, constant temperature).}$$

For example, the volume of the gas will decrease as the pressure increases. If you plot the relationship $PV = \text{constant}$ on a PV diagram, you find a hyperbola. [\[link\]](#) shows a graph of pressure versus volume. The hyperbolas represent ideal-gas behavior at various fixed temperatures, and are called *isotherms*. At lower temperatures, the curves begin to look less like hyperbolas—the gas is not behaving ideally and may even contain liquid. There is a **critical point**—that is, a **critical temperature**—above which liquid cannot exist. At sufficiently high pressure above the critical point, the gas will have the density of a liquid but will not condense. Carbon dioxide, for example, cannot be liquefied at a temperature above 31.0°C. **Critical pressure** is the minimum pressure needed for liquid to exist at the critical temperature. [\[link\]](#) lists representative critical temperatures and pressures.



PV diagrams. (a) Each curve (isotherm) represents the relationship

between P and V at a fixed temperature; the upper curves are at higher temperatures. The lower curves are not hyperbolas, because the gas is no longer an ideal gas. (b) An expanded portion of the PV diagram for low temperatures, where the phase can change from a gas to a liquid. The term “vapor” refers to the gas phase when it exists at a temperature below the boiling temperature.

Substance	Critical temperature		Critical pressure	
	K	°C	Pa	atm
Water	647.4	374.3	22.12×10^6	219.0
Sulfur dioxide	430.7	157.6	7.88×10^6	78.0
Ammonia	405.5	132.4	11.28×10^6	111.7
Carbon dioxide	304.2	31.1	7.39×10^6	73.2

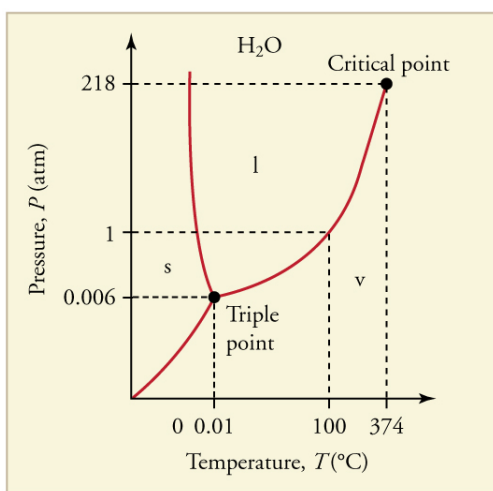
Substance	Critical temperature		Critical pressure	
	K	°C	Pa	atm
Oxygen	154.8	−118.4	5.08×10^6	50.3
Nitrogen	126.2	−146.9	3.39×10^6	33.6
Hydrogen	33.3	−239.9	1.30×10^6	12.9
Helium	5.3	−267.9	0.229×10^6	2.27

Critical Temperatures and Pressures

Phase Diagrams

The plots of pressure versus temperatures provide considerable insight into thermal properties of substances. There are well-defined regions on these graphs that correspond to various phases of matter, so PT graphs are called **phase diagrams**. [\[link\]](#) shows the phase diagram for water. Using the graph, if you know the pressure and temperature you can determine the phase of water. The solid lines—boundaries between phases—indicate temperatures and pressures at which the phases coexist (that is, they exist together in ratios, depending on pressure and temperature). For example, the boiling point of water is 100°C at 1.00 atm. As the pressure increases, the boiling temperature rises steadily to 374°C at a pressure of 218 atm. A pressure cooker (or even a covered pot) will cook food faster because the

water can exist as a liquid at temperatures greater than 100°C without all boiling away. The curve ends at a point called the *critical point*, because at higher temperatures the liquid phase does not exist at any pressure. The critical point occurs at the critical temperature, as you can see for water from [\[link\]](#). The critical temperature for oxygen is -118°C , so oxygen cannot be liquefied above this temperature.



The phase diagram (PT graph) for water. Note that the axes are nonlinear and the graph is not to scale. This graph is simplified—there are several other exotic phases of ice at higher pressures.

Similarly, the curve between the solid and liquid regions in [\[link\]](#) gives the melting temperature at various pressures. For example, the melting point is 0°C at 1.00 atm, as expected. Note that, at a fixed temperature, you can change the phase from solid (ice) to liquid (water) by increasing the pressure. Ice melts from pressure in the hands of a snowball maker. From

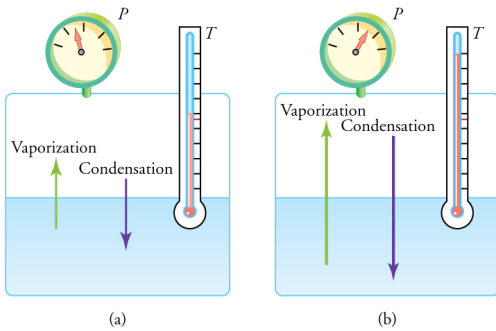
the phase diagram, we can also say that the melting temperature of ice rises with increased pressure. When a car is driven over snow, the increased pressure from the tires melts the snowflakes; afterwards the water refreezes and forms an ice layer.

At sufficiently low pressures there is no liquid phase, but the substance can exist as either gas or solid. For water, there is no liquid phase at pressures below 0.00600 atm. The phase change from solid to gas is called **sublimation**. It accounts for large losses of snow pack that never make it into a river, the routine automatic defrosting of a freezer, and the freeze-drying process applied to many foods. Carbon dioxide, on the other hand, sublimates at standard atmospheric pressure of 1 atm. (The solid form of CO_2 is known as dry ice because it does not melt. Instead, it moves directly from the solid to the gas state.)

All three curves on the phase diagram meet at a single point, the **triple point**, where all three phases exist in equilibrium. For water, the triple point occurs at 273.16 K (0.01°C), and is a more accurate calibration temperature than the melting point of water at 1.00 atm, or 273.15 K (0.0°C). See [\[link\]](#) for the triple point values of other substances.

Equilibrium

Liquid and gas phases are in equilibrium at the boiling temperature. (See [\[link\]](#).) If a substance is in a closed container at the boiling point, then the liquid is boiling and the gas is condensing at the same rate without net change in their relative amount. Molecules in the liquid escape as a gas at the same rate at which gas molecules stick to the liquid, or form droplets and become part of the liquid phase. The combination of temperature and pressure has to be “just right”; if the temperature and pressure are increased, equilibrium is maintained by the same increase of boiling and condensation rates.



Equilibrium between liquid and gas at two different boiling points inside a closed container.

(a) The rates of boiling and condensation are equal at this combination of temperature and pressure, so the liquid and gas phases are in equilibrium. (b) At a higher temperature, the boiling rate is faster and the rates at which molecules leave the liquid and enter the gas are also faster. Because there are more molecules in the gas, the gas pressure is higher and the rate at which gas molecules condense and enter the liquid is faster. As a result the gas and liquid are in equilibrium at this higher temperature.

Substance	Temperature		Pressure	
	K	°C	Pa	atm
Water	273.16	0.01	6.10×10^2	0.00600
Carbon dioxide	216.55	-56.60	5.16×10^5	5.11
Sulfur dioxide	197.68	-75.47	1.67×10^3	0.0167
Ammonia	195.40	-77.75	6.06×10^3	0.0600
Nitrogen	63.18	-210.0	1.25×10^4	0.124
Oxygen	54.36	-218.8	1.52×10^2	0.00151
Hydrogen	13.84	-259.3	7.04×10^3	0.0697

Triple Point Temperatures and Pressures

One example of equilibrium between liquid and gas is that of water and steam at 100°C and 1.00 atm. This temperature is the boiling point at that pressure, so they should exist in equilibrium. Why does an open pot of water at 100°C boil completely away? The gas surrounding an open pot is

not pure water: it is mixed with air. If pure water and steam are in a closed container at 100°C and 1.00 atm , they would coexist—but with air over the pot, there are fewer water molecules to condense, and water boils. What about water at 20.0°C and 1.00 atm ? This temperature and pressure correspond to the liquid region, yet an open glass of water at this temperature will completely evaporate. Again, the gas around it is air and not pure water vapor, so that the reduced evaporation rate is greater than the condensation rate of water from dry air. If the glass is sealed, then the liquid phase remains. We call the gas phase a **vapor** when it exists, as it does for water at 20.0°C , at a temperature below the boiling temperature.

Exercise:

Check Your Understanding

Problem:

Explain why a cup of water (or soda) with ice cubes stays at 0°C , even on a hot summer day.

Solution:

The ice and liquid water are in thermal equilibrium, so that the temperature stays at the freezing temperature as long as ice remains in the liquid. (Once all of the ice melts, the water temperature will start to rise.)

Vapor Pressure, Partial Pressure, and Dalton's Law

Vapor pressure is defined as the pressure at which a gas coexists with its solid or liquid phase. Vapor pressure is created by faster molecules that break away from the liquid or solid and enter the gas phase. The vapor pressure of a substance depends on both the substance and its temperature—an increase in temperature increases the vapor pressure.

Partial pressure is defined as the pressure a gas would create if it occupied the total volume available. In a mixture of gases, *the total pressure is the sum of partial pressures of the component gases*, assuming ideal gas behavior and no chemical reactions between the components. This law is

known as **Dalton's law of partial pressures**, after the English scientist John Dalton (1766–1844), who proposed it. Dalton's law is based on kinetic theory, where each gas creates its pressure by molecular collisions, independent of other gases present. It is consistent with the fact that pressures add according to [Pascal's Principle](#). Thus water evaporates and ice sublimates when their vapor pressures exceed the partial pressure of water vapor in the surrounding mixture of gases. If their vapor pressures are less than the partial pressure of water vapor in the surrounding gas, liquid droplets or ice crystals (frost) form.

Exercise:

Check Your Understanding

Problem:

Is energy transfer involved in a phase change? If so, will energy have to be supplied to change phase from solid to liquid and liquid to gas? What about gas to liquid and liquid to solid? Why do they spray the orange trees with water in Florida when the temperatures are near or just below freezing?

Solution:

Yes, energy transfer is involved in a phase change. We know that atoms and molecules in solids and liquids are bound to each other because we know that force is required to separate them. So in a phase change from solid to liquid and liquid to gas, a force must be exerted, perhaps by collision, to separate atoms and molecules. Force exerted through a distance is work, and energy is needed to do work to go from solid to liquid and liquid to gas. This is intuitively consistent with the need for energy to melt ice or boil water. The converse is also true. Going from gas to liquid or liquid to solid involves atoms and molecules pushing together, doing work and releasing energy.

Note:

PhET Explorations: States of Matter—Basics

Heat, cool, and compress atoms and molecules and watch as they change between solid, liquid, and gas phases.

https://phet.colorado.edu/sims/html/states-of-matter-basics/latest/states-of-matter-basics_en.html

Section Summary

- Most substances have three distinct phases: gas, liquid, and solid.
- Phase changes among the various phases of matter depend on temperature and pressure.
- The existence of the three phases with respect to pressure and temperature can be described in a phase diagram.
- Two phases coexist (i.e., they are in thermal equilibrium) at a set of pressures and temperatures. These are described as a line on a phase diagram.
- The three phases coexist at a single pressure and temperature. This is known as the triple point and is described by a single point on a phase diagram.
- A gas at a temperature below its boiling point is called a vapor.
- Vapor pressure is the pressure at which a gas coexists with its solid or liquid phase.
- Partial pressure is the pressure a gas would create if it existed alone.
- Dalton's law states that the total pressure is the sum of the partial pressures of all of the gases present.

Conceptual Questions

Exercise:

Problem:

A pressure cooker contains water and steam in equilibrium at a pressure greater than atmospheric pressure. How does this greater pressure increase cooking speed?

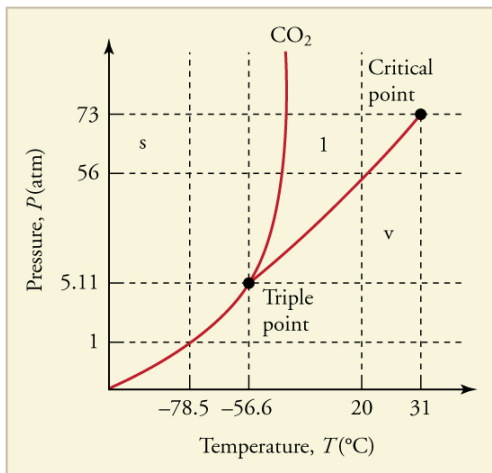
Exercise:

Problem:

Why does condensation form most rapidly on the coldest object in a room—for example, on a glass of ice water?

Exercise:**Problem:**

What is the vapor pressure of solid carbon dioxide (dry ice) at -78.5°C ?



The phase diagram for carbon dioxide. The axes are nonlinear, and the graph is not to scale. Dry ice is solid carbon dioxide and has a sublimation temperature of -78.5°C .

Exercise:**Problem:**

Can carbon dioxide be liquefied at room temperature (20°C)? If so, how? If not, why not? (See [\[link\]](#).)

Exercise:**Problem:**

Oxygen cannot be liquefied at room temperature by placing it under a large enough pressure to force its molecules together. Explain why this is.

Exercise:

Problem: What is the distinction between gas and vapor?

Glossary

PV diagram

a graph of pressure vs. volume

critical point

the temperature above which a liquid cannot exist

critical temperature

the temperature above which a liquid cannot exist

critical pressure

the minimum pressure needed for a liquid to exist at the critical temperature

vapor

a gas at a temperature below the boiling temperature

vapor pressure

the pressure at which a gas coexists with its solid or liquid phase

phase diagram

a graph of pressure vs. temperature of a particular substance, showing at which pressures and temperatures the three phases of the substance occur

triple point

the pressure and temperature at which a substance exists in equilibrium as a solid, liquid, and gas

sublimation

the phase change from solid to gas

partial pressure

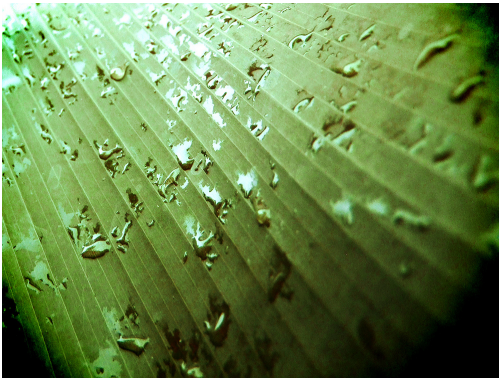
the pressure a gas would create if it occupied the total volume of space available

Dalton's law of partial pressures

the physical law that states that the total pressure of a gas is the sum of partial pressures of the component gases

Humidity, Evaporation, and Boiling

- Explain the relationship between vapor pressure of water and the capacity of air to hold water vapor.
- Explain the relationship between relative humidity and partial pressure of water vapor in the air.
- Calculate vapor density using vapor pressure.
- Calculate humidity and dew point.



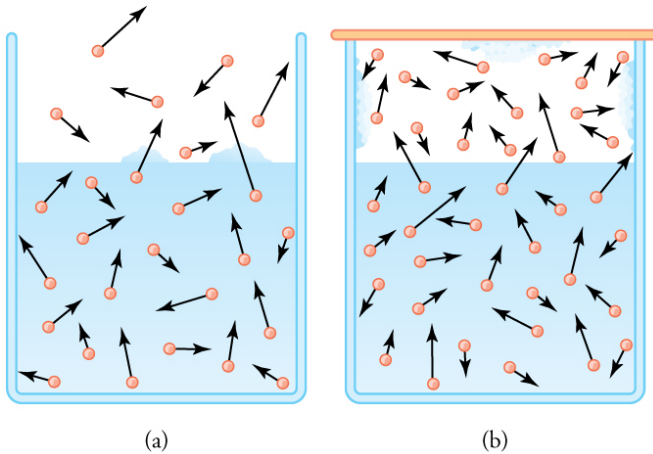
Dew drops like these, on a banana leaf photographed just after sunrise, form when the air temperature drops to or below the dew point. At the dew point, the rate at which water molecules join together is greater than the rate at which they separate, and some of the water condenses to form droplets. (credit: Aaron Escobar, Flickr)

The expression “it’s not the heat, it’s the humidity” makes a valid point. We keep cool in hot weather by evaporating sweat from our skin and water

from our breathing passages. Because evaporation is inhibited by high humidity, we feel hotter at a given temperature when the humidity is high. Low humidity, on the other hand, can cause discomfort from excessive drying of mucous membranes and can lead to an increased risk of respiratory infections.

When we say humidity, we really mean **relative humidity**. Relative humidity tells us how much water vapor is in the air compared with the maximum possible. At its maximum, denoted as **saturation**, the relative humidity is 100%, and evaporation is inhibited. The amount of water vapor in the air depends on temperature. For example, relative humidity rises in the evening, as air temperature declines, sometimes reaching the **dew point**. At the dew point temperature, relative humidity is 100%, and fog may result from the condensation of water droplets if they are small enough to stay in suspension. Conversely, if you wish to dry something (perhaps your hair), it is more effective to blow hot air over it rather than cold air, because, among other things, the increase in temperature increases the energy of the molecules, so the rate of evaporation increases.

The amount of water vapor in the air depends on the vapor pressure of water. The liquid and solid phases are continuously giving off vapor because some of the molecules have high enough speeds to enter the gas phase; see [\[link\]](#)(a). If a lid is placed over the container, as in [\[link\]](#)(b), evaporation continues, increasing the pressure, until sufficient vapor has built up for condensation to balance evaporation. Then equilibrium has been achieved, and the vapor pressure is equal to the partial pressure of water in the container. Vapor pressure increases with temperature because molecular speeds are higher as temperature increases. [\[link\]](#) gives representative values of water vapor pressure over a range of temperatures.



(a) Because of the distribution of speeds and kinetic energies, some water molecules can break away to the vapor phase even at temperatures below the ordinary boiling point. (b) If the container is sealed, evaporation will continue until there is enough vapor density for the condensation rate to equal the evaporation rate. This vapor density and the partial pressure it creates are the saturation values. They increase with temperature and are independent of the presence of other gases, such as air. They depend only on the vapor pressure of water.

Relative humidity is related to the partial pressure of water vapor in the air. At 100% humidity, the partial pressure is equal to the vapor pressure, and no more water can enter the vapor phase. If the partial pressure is less than the vapor pressure, then evaporation will take place, as humidity is less than 100%. If the partial pressure is greater than the vapor pressure, condensation takes place. In everyday language, people sometimes refer to

the capacity of air to “hold” water vapor, but this is not actually what happens. The water vapor is not held by the air. The amount of water in air is determined by the vapor pressure of water and has nothing to do with the properties of air.

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
−50	4.0	0.039
−20	1.04×10^2	0.89
−10	2.60×10^2	2.36
0	6.10×10^2	4.84
5	8.68×10^2	6.80
10	1.19×10^3	9.40

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
15	1.69×10^3	12.8
20	2.33×10^3	17.2
25	3.17×10^3	23.0
30	4.24×10^3	30.4
37	6.31×10^3	44.0
40	7.34×10^3	51.1
50	1.23×10^4	82.4
60	1.99×10^4	130
70	3.12×10^4	197

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
80	4.73×10^4	294
90	7.01×10^4	418
95	8.59×10^4	505
100	1.01×10^5	598
120	1.99×10^5	1095
150	4.76×10^5	2430
200	1.55×10^6	7090
220	2.32×10^6	10,200

Saturation Vapor Density of Water

Example:**Calculating Density Using Vapor Pressure**

[\[link\]](#) gives the vapor pressure of water at 20.0°C as 2.33×10^3 Pa. Use the ideal gas law to calculate the density of water vapor in g/m³ that would create a partial pressure equal to this vapor pressure. Compare the result with the saturation vapor density given in the table.

Strategy

To solve this problem, we need to break it down into a two steps. The partial pressure follows the ideal gas law,

Equation:

$$PV = nRT,$$

where n is the number of moles. If we solve this equation for n/V to calculate the number of moles per cubic meter, we can then convert this quantity to grams per cubic meter as requested. To do this, we need to use the molecular mass of water, which is given in the periodic table.

Solution

1. Identify the knowns and convert them to the proper units:

- a. temperature $T = 20^\circ\text{C} = 293$ K
- b. vapor pressure P of water at 20°C is 2.33×10^3 Pa
- c. molecular mass of water is 18.0 g/mol

2. Solve the ideal gas law for n/V .

Equation:

$$\frac{n}{V} = \frac{P}{RT}$$

3. Substitute known values into the equation and solve for n/V .

Equation:

$$\frac{n}{V} = \frac{P}{RT} = \frac{2.33 \times 10^3 \text{ Pa}}{(8.31 \text{ J/mol} \cdot \text{K})(293 \text{ K})} = 0.957 \text{ mol/m}^3$$

4. Convert the density in moles per cubic meter to grams per cubic meter.

Equation:

$$\rho = \left(0.957 \frac{\text{mol}}{\text{m}^3}\right) \left(\frac{18.0 \text{ g}}{\text{mol}}\right) = 17.2 \text{ g/m}^3$$

Discussion

The density is obtained by assuming a pressure equal to the vapor pressure of water at 20.0°C. The density found is identical to the value in [\[link\]](#), which means that a vapor density of 17.2 g/m³ at 20.0°C creates a partial pressure of 2.33×10^3 Pa, equal to the vapor pressure of water at that temperature. If the partial pressure is equal to the vapor pressure, then the liquid and vapor phases are in equilibrium, and the relative humidity is 100%. Thus, there can be no more than 17.2 g of water vapor per m³ at 20.0°C, so that this value is the saturation vapor density at that temperature. This example illustrates how water vapor behaves like an ideal gas: the pressure and density are consistent with the ideal gas law (assuming the density in the table is correct). The saturation vapor densities listed in [\[link\]](#) are the maximum amounts of water vapor that air can hold at various temperatures.

Note:**Percent Relative Humidity**

We define **percent relative humidity** as the ratio of vapor density to saturation vapor density, or

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100$$

We can use this and the data in [\[link\]](#) to do a variety of interesting calculations, keeping in mind that relative humidity is based on the comparison of the partial pressure of water vapor in air and ice.

Example:**Calculating Humidity and Dew Point**

(a) Calculate the percent relative humidity on a day when the temperature is 25.0°C and the air contains 9.40 g of water vapor per m³. (b) At what temperature will this air reach 100% relative humidity (the saturation density)? This temperature is the dew point. (c) What is the humidity when the air temperature is 25.0°C and the dew point is – 10.0°C?

Strategy and Solution

(a) Percent relative humidity is defined as the ratio of vapor density to saturation vapor density.

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100$$

The first is given to be 9.40 g/m³, and the second is found in [\[link\]](#) to be 23.0 g/m³. Thus,

Equation:

$$\text{percent relative humidity} = \frac{9.40 \text{ g/m}^3}{23.0 \text{ g/m}^3} \times 100 = 40.9\%$$

(b) The air contains 9.40 g/m³ of water vapor. The relative humidity will be 100% at a temperature where 9.40 g/m³ is the saturation density. Inspection of [\[link\]](#) reveals this to be the case at 10.0°C, where the relative humidity will be 100%. That temperature is called the dew point for air with this concentration of water vapor.

(c) Here, the dew point temperature is given to be – 10.0°C. Using [\[link\]](#), we see that the vapor density is 2.36 g/m³, because this value is the saturation vapor density at – 10.0°C. The saturation vapor density at 25.0°C is seen to be 23.0 g/m³. Thus, the relative humidity at 25.0°C is

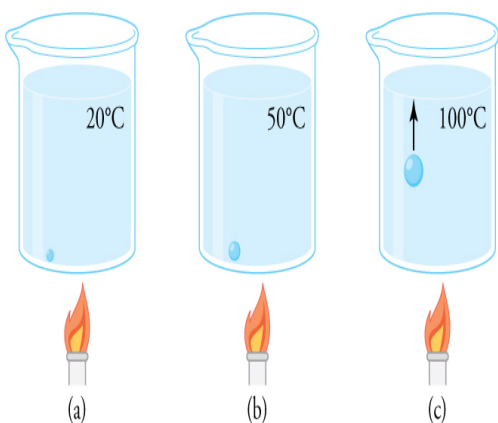
Equation:

$$\text{percent relative humidity} = \frac{2.36 \text{ g/m}^3}{23.0 \text{ g/m}^3} \times 100 = 10.3\%.$$

Discussion

The importance of dew point is that air temperature cannot drop below 10.0°C in part (b), or -10.0°C in part (c), without water vapor condensing out of the air. If condensation occurs, considerable transfer of heat occurs (discussed in [Heat and Heat Transfer Methods](#)), which prevents the temperature from further dropping. When dew points are below 0°C , freezing temperatures are a greater possibility, which explains why farmers keep track of the dew point. Low humidity in deserts means low dew-point temperatures. Thus condensation is unlikely. If the temperature drops, vapor does not condense in liquid drops. Because no heat is released into the air, the air temperature drops more rapidly compared to air with higher humidity. Likewise, at high temperatures, liquid droplets do not evaporate, so that no heat is removed from the gas to the liquid phase. This explains the large range of temperature in arid regions.

Why does water boil at 100°C ? You will note from [\[link\]](#) that the vapor pressure of water at 100°C is $1.01 \times 10^5 \text{ Pa}$, or 1.00 atm. Thus, it can evaporate without limit at this temperature and pressure. But why does it form bubbles when it boils? This is because water ordinarily contains significant amounts of dissolved air and other impurities, which are observed as small bubbles of air in a glass of water. If a bubble starts out at the bottom of the container at 20°C , it contains water vapor (about 2.30%). The pressure inside the bubble is fixed at 1.00 atm (we ignore the slight pressure exerted by the water around it). As the temperature rises, the amount of air in the bubble stays the same, but the water vapor increases; the bubble expands to keep the pressure at 1.00 atm. At 100°C , water vapor enters the bubble continuously since the partial pressure of water is equal to 1.00 atm in equilibrium. It cannot reach this pressure, however, since the bubble also contains air and total pressure is 1.00 atm. The bubble grows in size and thereby increases the buoyant force. The bubble breaks away and rises rapidly to the surface—we call this boiling! (See [\[link\]](#).)



(a) An air bubble in water starts out saturated with water vapor at 20°C. (b) As the temperature rises, water vapor enters the bubble because its vapor pressure increases. The bubble expands to keep its pressure at 1.00 atm.

(c) At 100°C, water vapor enters the bubble continuously because water's vapor pressure exceeds its partial pressure in the bubble, which must be less than 1.00 atm. The bubble grows and rises to the surface.

Exercise:
Check Your Understanding

Problem:

Freeze drying is a process in which substances, such as foods, are dried by placing them in a vacuum chamber and lowering the atmospheric pressure around them. How does the lowered atmospheric pressure speed the drying process, and why does it cause the temperature of the food to drop?

Solution:

Decreased the atmospheric pressure results in decreased partial pressure of water, hence a lower humidity. So evaporation of water from food, for example, will be enhanced. The molecules of water most likely to break away from the food will be those with the greatest velocities. Those remaining thus have a lower average velocity and a lower temperature. This can (and does) result in the freezing and drying of the food; hence the process is aptly named freeze drying.

Note:

PhET Explorations: States of Matter

Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.

https://phet.colorado.edu/sims/html/states-of-matter/latest/states-of-matter_en.html

Section Summary

- Relative humidity is the fraction of water vapor in a gas compared to the saturation value.
- The saturation vapor density can be determined from the vapor pressure for a given temperature.

- Percent relative humidity is defined to be

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100.$$

- The dew point is the temperature at which air reaches 100% relative humidity.

Conceptual Questions

Exercise:

Problem:

Because humidity depends only on water's vapor pressure and temperature, are the saturation vapor densities listed in [\[link\]](#) valid in an atmosphere of helium at a pressure of $1.01 \times 10^5 \text{ N/m}^2$, rather than air? Are those values affected by altitude on Earth?

Exercise:

Problem:

Why does a beaker of 40.0°C water placed in a vacuum chamber start to boil as the chamber is evacuated (air is pumped out of the chamber)? At what pressure does the boiling begin? Would food cook any faster in such a beaker?

Exercise:

Problem:

Why does rubbing alcohol evaporate much more rapidly than water at STP (standard temperature and pressure)?

Problems & Exercises

Exercise:

Problem:

Dry air is 78.1% nitrogen. What is the partial pressure of nitrogen when the atmospheric pressure is $1.01 \times 10^5 \text{ N/m}^2$?

Solution:

$$7.89 \times 10^4 \text{ Pa}$$

Exercise:**Problem:**

(a) What is the vapor pressure of water at 20.0°C? (b) What percentage of atmospheric pressure does this correspond to? (c) What percent of 20.0°C air is water vapor if it has 100% relative humidity? (The density of dry air at 20.0°C is 1.20 kg/m^3 .)

Exercise:**Problem:**

Pressure cookers increase cooking speed by raising the boiling temperature of water above its value at atmospheric pressure. (a) What pressure is necessary to raise the boiling point to 120.0°C? (b) What gauge pressure does this correspond to?

Solution:

(a) $1.99 \times 10^5 \text{ Pa}$

(b) 0.97 atm

Exercise:

Problem:

(a) At what temperature does water boil at an altitude of 1500 m (about 5000 ft) on a day when atmospheric pressure is $8.59 \times 10^4 \text{ N/m}^2$? (b) What about at an altitude of 3000 m (about 10,000 ft) when atmospheric pressure is $7.00 \times 10^4 \text{ N/m}^2$?

Exercise:**Problem:**

What is the atmospheric pressure on top of Mt. Everest on a day when water boils there at a temperature of 70.0°C ?

Solution:

$$3.12 \times 10^4 \text{ Pa}$$

Exercise:**Problem:**

At a spot in the high Andes, water boils at 80.0°C , greatly reducing the cooking speed of potatoes, for example. What is atmospheric pressure at this location?

Exercise:**Problem:**

What is the relative humidity on a 25.0°C day when the air contains 18.0 g/m^3 of water vapor?

Solution:

$$78.3\%$$

Exercise:

Problem:

What is the density of water vapor in g/m^3 on a hot dry day in the desert when the temperature is 40.0°C and the relative humidity is 6.00%?

Exercise:**Problem:**

A deep-sea diver should breathe a gas mixture that has the same oxygen partial pressure as at sea level, where dry air contains 20.9% oxygen and has a total pressure of $1.01 \times 10^5 \text{ N}/\text{m}^2$. (a) What is the partial pressure of oxygen at sea level? (b) If the diver breathes a gas mixture at a pressure of $2.00 \times 10^6 \text{ N}/\text{m}^2$, what percent oxygen should it be to have the same oxygen partial pressure as at sea level?

Solution:

(a) $2.12 \times 10^4 \text{ Pa}$

(b) 1.06 %

Exercise:**Problem:**

The vapor pressure of water at 40.0°C is $7.34 \times 10^3 \text{ N}/\text{m}^2$. Using the ideal gas law, calculate the density of water vapor in g/m^3 that creates a partial pressure equal to this vapor pressure. The result should be the same as the saturation vapor density at that temperature ($51.1 \text{ g}/\text{m}^3$).

Exercise:

Problem:

Air in human lungs has a temperature of 37.0°C and a saturation vapor density of 44.0 g/m^3 . (a) If 2.00 L of air is exhaled and very dry air inhaled, what is the maximum loss of water vapor by the person? (b) Calculate the partial pressure of water vapor having this density, and compare it with the vapor pressure of $6.31 \times 10^3\text{ N/m}^2$.

Solution:

(a) $8.80 \times 10^{-2}\text{ g}$

(b) $6.30 \times 10^3\text{ Pa}$; the two values are nearly identical.

Exercise:**Problem:**

If the relative humidity is 90.0% on a muggy summer morning when the temperature is 20.0°C , what will it be later in the day when the temperature is 30.0°C , assuming the water vapor density remains constant?

Exercise:**Problem:**

Late on an autumn day, the relative humidity is 45.0% and the temperature is 20.0°C . What will the relative humidity be that evening when the temperature has dropped to 10.0°C , assuming constant water vapor density?

Solution:

82.3%

Exercise:

Problem:

Atmospheric pressure atop Mt. Everest is $3.30 \times 10^4 \text{ N/m}^2$. (a) What is the partial pressure of oxygen there if it is 20.9% of the air? (b) What percent oxygen should a mountain climber breathe so that its partial pressure is the same as at sea level, where atmospheric pressure is $1.01 \times 10^5 \text{ N/m}^2$? (c) One of the most severe problems for those climbing very high mountains is the extreme drying of breathing passages. Why does this drying occur?

Exercise:**Problem:**

What is the dew point (the temperature at which 100% relative humidity would occur) on a day when relative humidity is 39.0% at a temperature of 20.0°C ?

Solution:

4.77°C

Exercise:**Problem:**

On a certain day, the temperature is 25.0°C and the relative humidity is 90.0%. How many grams of water must condense out of each cubic meter of air if the temperature falls to 15.0°C ? Such a drop in temperature can, thus, produce heavy dew or fog.

Exercise:**Problem: Integrated Concepts**

The boiling point of water increases with depth because pressure increases with depth. At what depth will fresh water have a boiling point of 150°C , if the surface of the water is at sea level?

Solution:

38.3 m

Exercise:**Problem: Integrated Concepts**

(a) At what depth in fresh water is the critical pressure of water reached, given that the surface is at sea level? (b) At what temperature will this water boil? (c) Is a significantly higher temperature needed to boil water at a greater depth?

Exercise:**Problem: Integrated Concepts**

To get an idea of the small effect that temperature has on Archimedes' principle, calculate the fraction of a copper block's weight that is supported by the buoyant force in 0°C water and compare this fraction with the fraction supported in 95.0°C water.

Solution:

$\frac{(F_B/w_{Cu})}{(F_B/w_{Cu})'} = 1.02$. The buoyant force supports nearly the exact same amount of force on the copper block in both circumstances.

Exercise:**Problem: Integrated Concepts**

If you want to cook in water at 150°C, you need a pressure cooker that can withstand the necessary pressure. (a) What pressure is required for the boiling point of water to be this high? (b) If the lid of the pressure cooker is a disk 25.0 cm in diameter, what force must it be able to withstand at this pressure?

Exercise:

Problem: Unreasonable Results

(a) How many moles per cubic meter of an ideal gas are there at a pressure of $1.00 \times 10^{14} \text{ N/m}^2$ and at 0°C ? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) $4.41 \times 10^{10} \text{ mol/m}^3$

(b) It's unreasonably large.

(c) At high pressures such as these, the ideal gas law can no longer be applied. As a result, unreasonable answers come up when it is used.

Exercise:**Problem: Unreasonable Results**

(a) An automobile mechanic claims that an aluminum rod fits loosely into its hole on an aluminum engine block because the engine is hot and the rod is cold. If the hole is 10.0% bigger in diameter than the 22.0°C rod, at what temperature will the rod be the same size as the hole? (b) What is unreasonable about this temperature? (c) Which premise is responsible?

Exercise:**Problem: Unreasonable Results**

The temperature inside a supernova explosion is said to be $2.00 \times 10^{13} \text{ K}$. (a) What would the average velocity v_{rms} of hydrogen atoms be? (b) What is unreasonable about this velocity? (c) Which premise or assumption is responsible?

Solution:

(a) $7.03 \times 10^8 \text{ m/s}$

(b) The velocity is too high—it's greater than the speed of light.

(c) The assumption that hydrogen inside a supernova behaves as an idea gas is responsible, because of the great temperature and density in the core of a star. Furthermore, when a velocity greater than the speed of light is obtained, classical physics must be replaced by relativity, a subject not yet covered.

Exercise:

Problem: Unreasonable Results

Suppose the relative humidity is 80% on a day when the temperature is 30.0°C . (a) What will the relative humidity be if the air cools to 25.0°C and the vapor density remains constant? (b) What is unreasonable about this result? (c) Which premise is responsible?

Glossary

dew point

the temperature at which relative humidity is 100%; the temperature at which water starts to condense out of the air

saturation

the condition of 100% relative humidity

percent relative humidity

the ratio of vapor density to saturation vapor density

relative humidity

the amount of water in the air relative to the maximum amount the air can hold

Introduction to Heat and Heat Transfer Methods

class="introduction"

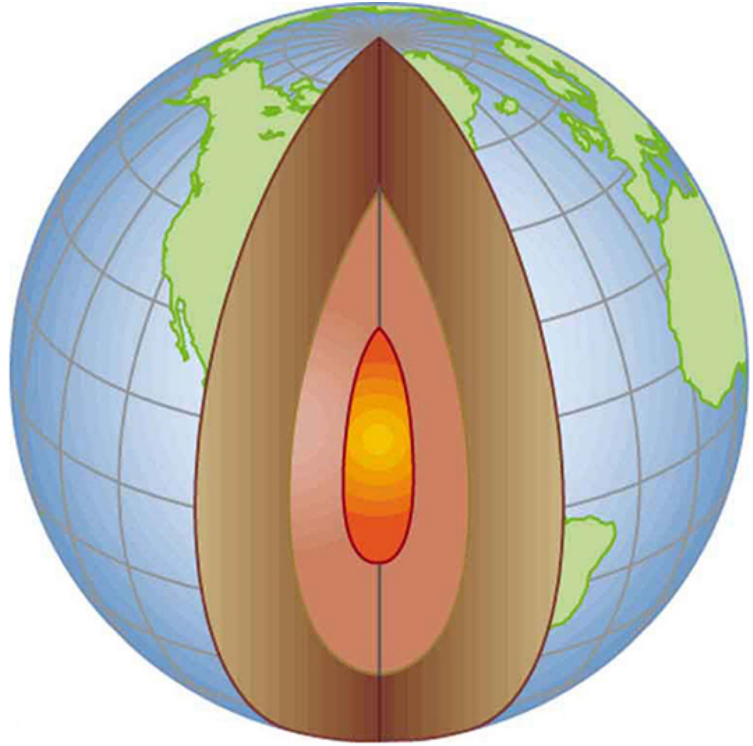
(a) The chilling effect of a clear breezy night is produced by the wind and by radiative heat transfer to cold outer space. (b) There was once great controversy about the Earth's age, but it is now generally accepted to be about 4.5 billion years old. Much of the debate is centered on the Earth's molten interior. According to our understanding of heat transfer, if the Earth is really that old, its

center should
have cooled
off long ago.

The
discovery of
radioactivity
in rocks
revealed the
source of
energy that
keeps the
Earth's
interior
molten,
despite heat
transfer to the
surface, and
from there to
cold outer
space.



(a)



(b)

Energy can exist in many forms and heat is one of the most intriguing. Heat is often hidden, as it only exists when in transit, and is transferred by a number of distinctly different methods. Heat transfer touches every aspect of our lives and helps us understand how the universe functions. It explains the chill we feel on a clear breezy night, or why Earth's core has yet to cool. This chapter defines and explores heat transfer, its effects, and the methods by which heat is transferred. These topics are fundamental, as well as practical, and will often be referred to in the chapters ahead.

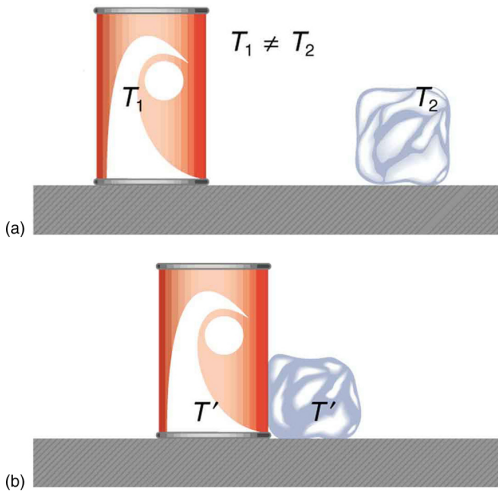
Heat

- Define heat as transfer of energy.

In [Work, Energy, and Energy Resources](#), we defined work as force times distance and learned that work done on an object changes its kinetic energy. We also saw in [Temperature, Kinetic Theory, and the Gas Laws](#) that temperature is proportional to the (average) kinetic energy of atoms and molecules. We say that a thermal system has a certain internal energy: its internal energy is higher if the temperature is higher. If two objects at different temperatures are brought in contact with each other, energy is transferred from the hotter to the colder object until equilibrium is reached and the bodies reach thermal equilibrium (i.e., they are at the same temperature). No work is done by either object, because no force acts through a distance. The transfer of energy is caused by the temperature difference, and ceases once the temperatures are equal. These observations lead to the following definition of **heat**: Heat is the spontaneous transfer of energy due to a temperature difference.

As noted in [Temperature, Kinetic Theory, and the Gas Laws](#), heat is often confused with temperature. For example, we may say the heat was unbearable, when we actually mean that the temperature was high. Heat is a form of energy, whereas temperature is not. The misconception arises because we are sensitive to the flow of heat, rather than the temperature.

Owing to the fact that heat is a form of energy, it has the SI unit of *joule* (J). The *calorie* (cal) is a common unit of energy, defined as the energy needed to change the temperature of 1.00 g of water by 1.00°C—specifically, between 14.5°C and 15.5°C, since there is a slight temperature dependence. Perhaps the most common unit of heat is the **kilocalorie** (kcal), which is the energy needed to change the temperature of 1.00 kg of water by 1.00°C. Since mass is most often specified in kilograms, kilocalorie is commonly used. Food calories (given the notation Cal, and sometimes called “big calorie”) are actually kilocalories (1 kilocalorie = 1000 calories), a fact not easily determined from package labeling.



In figure (a) the soft drink and the ice have different temperatures, T_1 and T_2 , and are not in thermal equilibrium. In figure (b), when the soft drink and ice are allowed to interact, energy is transferred until they reach the same temperature T' , achieving equilibrium. Heat transfer occurs due to the difference in temperatures. In fact, since the soft drink and ice are both in contact with the surrounding air and bench, the equilibrium temperature will be the same for both.

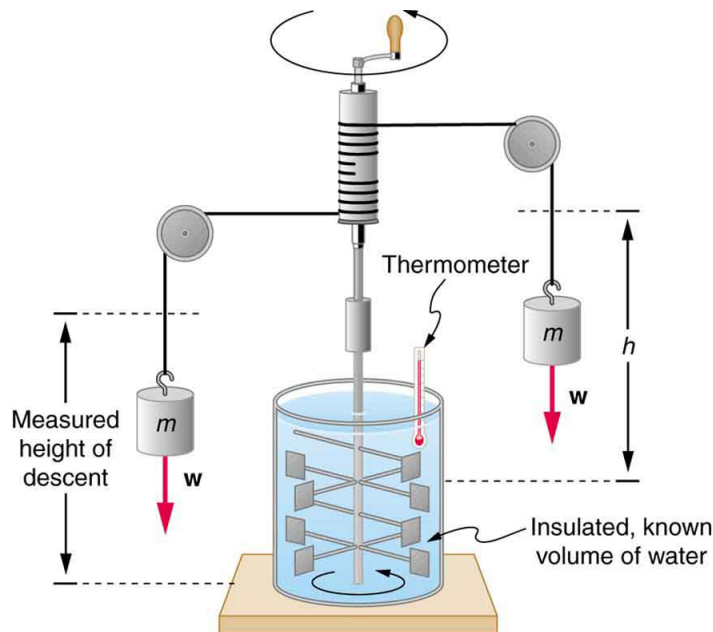
Mechanical Equivalent of Heat

It is also possible to change the temperature of a substance by doing work. Work can transfer energy into or out of a system. This realization helped establish the fact that heat is a form of energy. James Prescott Joule (1818–1889) performed many experiments to establish the **mechanical equivalent of heat**—*the work needed to produce the same effects as heat transfer*. In terms of the units used for these two terms, the best modern value for this equivalence is

Equation:

$$1.000 \text{ kcal} = 4186 \text{ J.}$$

We consider this equation as the conversion between two different units of energy.



Schematic depiction of Joule's experiment that established the equivalence of heat and work.

The figure above shows one of Joule's most famous experimental setups for demonstrating the mechanical equivalent of heat. It demonstrated that work and heat can produce the same effects, and helped establish the principle of conservation of energy. Gravitational potential energy (PE) (work done by the gravitational force) is converted into kinetic energy (KE), and then randomized by viscosity and turbulence into increased average kinetic energy of atoms and molecules in the system, producing a temperature increase. His contributions to the field of thermodynamics were so significant that the SI unit of energy was named after him.

Heat added or removed from a system changes its internal energy and thus its temperature. Such a temperature increase is observed while cooking. However, adding heat does not necessarily increase the temperature. An example is melting of ice; that is, when a substance changes from one phase to another. Work done on the system or by the system can also change the internal energy of the system. Joule demonstrated that the temperature of a system can be increased by stirring. If an ice cube is rubbed against a rough surface, work is done by the frictional force. A system has a well-defined internal energy, but we cannot say that it has a certain "heat content" or "work content". We use the phrase "heat transfer" to emphasize its nature.

Exercise:

Check Your Understanding

Problem:

Two samples (A and B) of the same substance are kept in a lab. Someone adds 10 kilojoules (kJ) of heat to one sample, while 10 kJ of work is done on the other sample. How can you tell to which sample the heat was added?

Solution:

Heat and work both change the internal energy of the substance. However, the properties of the sample only depend on the internal energy so that it is impossible to tell whether heat was added to sample A or B.

Summary

- Heat and work are the two distinct methods of energy transfer.
- Heat is energy transferred solely due to a temperature difference.
- Any energy unit can be used for heat transfer, and the most common are kilocalorie (kcal) and joule (J).
- Kilocalorie is defined to be the energy needed to change the temperature of 1.00 kg of water between 14.5°C and 15.5°C.
- The mechanical equivalent of this heat transfer is
 $1.00 \text{ kcal} = 4186 \text{ J}$.

Conceptual Questions

Exercise:

Problem: How is heat transfer related to temperature?

Exercise:

Problem:

Describe a situation in which heat transfer occurs. What are the resulting forms of energy?

Exercise:

Problem:

When heat transfers into a system, is the energy stored as heat? Explain briefly.

Glossary

heat

the spontaneous transfer of energy due to a temperature difference

kilocalorie

1 kilocalorie = 1000 calories

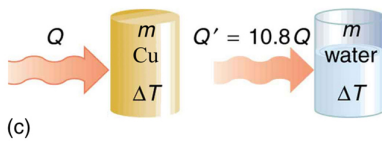
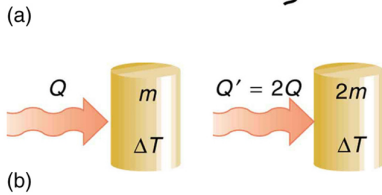
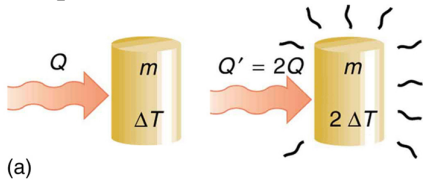
mechanical equivalent of heat

the work needed to produce the same effects as heat transfer

Temperature Change and Heat Capacity

- Observe heat transfer and change in temperature and mass.
- Calculate final temperature after heat transfer between two objects.

One of the major effects of heat transfer is temperature change: heating increases the temperature while cooling decreases it. We assume that there is no phase change and that no work is done on or by the system. Experiments show that the transferred heat depends on three factors—the change in temperature, the mass of the system, and the substance and phase of the substance.



The heat Q transferred to cause a temperature change depends on the magnitude of the temperature change, the mass of the system, and the substance and phase involved. (a) The amount of heat transferred is directly proportional to the temperature change.

To double the temperature change of a mass m , you need to add twice the heat. (b) The amount of heat transferred is also directly proportional to the mass.

To cause an equivalent temperature change in a

doubled mass, you need to add twice the heat. (c)

The amount of heat transferred depends on the substance and its phase. If it takes an amount Q of heat to cause a temperature change ΔT in a given mass of copper, it will take 10.8 times that amount of heat to cause the equivalent temperature change in the same mass of water assuming no phase change in either substance.

The dependence on temperature change and mass are easily understood. Owing to the fact that the (average) kinetic energy of an atom or molecule is proportional to the absolute temperature, the internal energy of a system is proportional to the absolute temperature and the number of atoms or molecules. Owing to the fact that the transferred heat is equal to the change in the internal energy, the heat is proportional to the mass of the substance and the temperature change. The transferred heat also depends on the substance so that, for example, the heat necessary to raise the temperature is less for alcohol than for water. For the same substance, the transferred heat also depends on the phase (gas, liquid, or solid).

Note:

Heat Transfer and Temperature Change

The quantitative relationship between heat transfer and temperature change contains all three factors:

Equation:

$$Q = mc\Delta T,$$

where Q is the symbol for heat transfer, m is the mass of the substance, and ΔT is the change in temperature. The symbol c stands for **specific heat** and depends on the material and phase. The specific heat is the amount of heat necessary to change the

temperature of 1.00 kg of mass by 1.00°C. The specific heat c is a property of the substance; its SI unit is J/(kg · K) or J/(kg · °C). Recall that the temperature change (ΔT) is the same in units of kelvin and degrees Celsius. If heat transfer is measured in kilocalories, then *the unit of specific heat* is kcal/(kg · °C).

Values of specific heat must generally be looked up in tables, because there is no simple way to calculate them. In general, the specific heat also depends on the temperature. [\[link\]](#) lists representative values of specific heat for various substances. Except for gases, the temperature and volume dependence of the specific heat of most substances is weak. We see from this table that the specific heat of water is five times that of glass and ten times that of iron, which means that it takes five times as much heat to raise the temperature of water the same amount as for glass and ten times as much heat to raise the temperature of water as for iron. In fact, water has one of the largest specific heats of any material, which is important for sustaining life on Earth.

Example:

Calculating the Required Heat: Heating Water in an Aluminum Pan

A 0.500 kg aluminum pan on a stove is used to heat 0.250 liters of water from 20.0°C to 80.0°C. (a) How much heat is required? What percentage of the heat is used to raise the temperature of (b) the pan and (c) the water?

Strategy

The pan and the water are always at the same temperature. When you put the pan on the stove, the temperature of the water and the pan is increased by the same amount. We use the equation for the heat transfer for the given temperature change and mass of water and aluminum. The specific heat values for water and aluminum are given in [\[link\]](#).

Solution

Because water is in thermal contact with the aluminum, the pan and the water are at the same temperature.

1. Calculate the temperature difference:

Equation:

$$\Delta T = T_f - T_i = 60.0^\circ\text{C}.$$

2. Calculate the mass of water. Because the density of water is 1000 kg/m³, one liter of water has a mass of 1 kg, and the mass of 0.250 liters of water is $m_w = 0.250$ kg.
3. Calculate the heat transferred to the water. Use the specific heat of water in [\[link\]](#):

Equation:

$$Q_w = m_w c_w \Delta T = (0.250 \text{ kg})(4186 \text{ J/kg}^\circ\text{C})(60.0^\circ\text{C}) = 62.8 \text{ kJ}.$$

4. Calculate the heat transferred to the aluminum. Use the specific heat for aluminum in [\[link\]](#):

Equation:

$$Q_{Al} = m_{Al} c_{Al} \Delta T = (0.500 \text{ kg})(900 \text{ J/kg}^\circ\text{C})(60.0^\circ\text{C}) = 27.0 \times 10^4 \text{ J} = 27.0 \text{ kJ}.$$

5. Compare the percentage of heat going into the pan versus that going into the water. First, find the total transferred heat:

Equation:

$$Q_{\text{Total}} = Q_W + Q_{Al} = 62.8 \text{ kJ} + 27.0 \text{ kJ} = 89.8 \text{ kJ}.$$

Thus, the amount of heat going into heating the pan is

Equation:

$$\frac{27.0 \text{ kJ}}{89.8 \text{ kJ}} \times 100\% = 30.1\%,$$

and the amount going into heating the water is

Equation:

$$\frac{62.8 \text{ kJ}}{89.8 \text{ kJ}} \times 100\% = 69.9\%.$$

Discussion

In this example, the heat transferred to the container is a significant fraction of the total transferred heat. Although the mass of the pan is twice that of the water, the specific heat of water is over four times greater than that of aluminum. Therefore, it takes a bit more than twice the heat to achieve the given temperature change for the water as compared to the aluminum pan.



The smoking brakes on this truck are a visible evidence of the mechanical equivalent of heat.

Example:

Calculating the Temperature Increase from the Work Done on a Substance: Truck Brakes Overheat on Downhill Runs

Truck brakes used to control speed on a downhill run do work, converting gravitational potential energy into increased internal energy (higher temperature) of the brake material. This conversion prevents the gravitational potential energy from being converted into kinetic energy of the truck. The problem is that the mass of the truck is large compared with that of the brake material absorbing the energy, and the temperature increase may occur too fast for sufficient heat to transfer from the brakes to the environment.

Calculate the temperature increase of 100 kg of brake material with an average specific heat of $800 \text{ J/kg} \cdot ^\circ\text{C}$ if the material retains 10% of the energy from a 10,000-kg truck descending 75.0 m (in vertical displacement) at a constant speed.

Strategy

If the brakes are not applied, gravitational potential energy is converted into kinetic energy. When brakes are applied, gravitational potential energy is converted into internal energy of the brake material. We first calculate the gravitational potential energy (Mgh) that the entire truck loses in its descent and then find the temperature increase produced in the brake material alone.

Solution

1. Calculate the change in gravitational potential energy as the truck goes downhill

Equation:

$$Mgh = (10,000 \text{ kg}) (9.80 \text{ m/s}^2) (75.0 \text{ m}) = 7.35 \times 10^6 \text{ J.}$$

2. Calculate the temperature from the heat transferred using $Q=Mgh$ and

Equation:

$$\Delta T = \frac{Q}{mc},$$

where m is the mass of the brake material. Insert the values $m = 100 \text{ kg}$ and $c = 800 \text{ J/kg} \cdot ^\circ\text{C}$ to find

Equation:

$$\Delta T = \frac{(7.35 \times 10^5 \text{ J})}{(100 \text{ kg})(800 \text{ J/kg}^\circ\text{C})} = 9.2^\circ\text{C.}$$

Discussion

This same idea underlies the recent hybrid technology of cars, where mechanical energy (gravitational potential energy) is converted by the brakes into electrical energy (battery).

Substances	Specific heat (c)	
Solids	J/kg·°C	kcal/kg·°C [footnote] These values are identical in units of cal/g ·°C.
Aluminum	900	0.215
Asbestos	800	0.19
Concrete, granite (average)	840	0.20
Copper	387	0.0924
Glass	840	0.20

Substances	Specific heat (c)	
Gold	129	0.0308
Human body (average at 37 °C)	3500	0.83
Ice (average, -50°C to 0°C)	2090	0.50
Iron, steel	452	0.108
Lead	128	0.0305
Silver	235	0.0562
Wood	1700	0.4
<i>Liquids</i>		
Benzene	1740	0.415
Ethanol	2450	0.586
Glycerin	2410	0.576
Mercury	139	0.0333
Water (15.0 °C)	4186	1.000
<i>Gases</i> [footnote] c_v at constant volume and at 20.0°C, except as noted, and at 1.00 atm average pressure. Values in parentheses are c_p at a constant pressure of 1.00 atm.		
Air (dry)	721 (1015)	0.172 (0.242)
Ammonia	1670 (2190)	0.399 (0.523)
Carbon dioxide	638 (833)	0.152 (0.199)

Substances	Specific heat (c)	
Nitrogen	739 (1040)	0.177 (0.248)
Oxygen	651 (913)	0.156 (0.218)
Steam (100°C)	1520 (2020)	0.363 (0.482)

Specific Heats^[footnote] of Various Substances

The values for solids and liquids are at constant volume and at 25°C, except as noted.

Note that [\[link\]](#) is an illustration of the mechanical equivalent of heat. Alternatively, the temperature increase could be produced by a blow torch instead of mechanically.

Example:

Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan

Suppose you pour 0.250 kg of 20.0°C water (about a cup) into a 0.500-kg aluminum pan off the stove with a temperature of 150°C. Assume that the pan is placed on an insulated pad and that a negligible amount of water boils off. What is the temperature when the water and pan reach thermal equilibrium a short time later?

Strategy

The pan is placed on an insulated pad so that little heat transfer occurs with the surroundings. Originally the pan and water are not in thermal equilibrium: the pan is at a higher temperature than the water. Heat transfer then restores thermal equilibrium once the water and pan are in contact. Because heat transfer between the pan and water takes place rapidly, the mass of evaporated water is negligible and the magnitude of the heat lost by the pan is equal to the heat gained by the water. The exchange of heat stops once a thermal equilibrium between the pan and the water is achieved. The heat exchange can be written as $|Q_{\text{hot}}| = Q_{\text{cold}}$.

Solution

1. Use the equation for heat transfer $Q = mc\Delta T$ to express the heat lost by the aluminum pan in terms of the mass of the pan, the specific heat of aluminum, the initial temperature of the pan, and the final temperature:

Equation:

$$Q_{\text{hot}} = m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150^{\circ}\text{C}).$$

2. Express the heat gained by the water in terms of the mass of the water, the specific heat of water, the initial temperature of the water and the final temperature:

Equation:

$$Q_{\text{cold}} = m_{\text{W}}c_{\text{W}}(T_{\text{f}} - 20.0^{\circ}\text{C}).$$

3. Note that $Q_{\text{hot}} < 0$ and $Q_{\text{cold}} > 0$ and that they must sum to zero because the heat lost by the hot pan must be the same as the heat gained by the cold water:

Equation:

$$\begin{aligned}Q_{\text{cold}} + Q_{\text{hot}} &= 0, \\Q_{\text{cold}} &= -Q_{\text{hot}}, \\m_{\text{W}}c_{\text{W}}(T_{\text{f}} - 20.0^{\circ}\text{C}) &= -m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150^{\circ}\text{C}).\end{aligned}$$

4. This an equation for the unknown final temperature, T_{f}
5. Bring all terms involving T_{f} on the left hand side and all other terms on the right hand side. Solve for T_{f} ,

Equation:

$$T_{\text{f}} = \frac{m_{\text{Al}}c_{\text{Al}}(150^{\circ}\text{C}) + m_{\text{W}}c_{\text{W}}(20.0^{\circ}\text{C})}{m_{\text{Al}}c_{\text{Al}} + m_{\text{W}}c_{\text{W}}},$$

and insert the numerical values:

Equation:

$$\begin{aligned}T_{\text{f}} &= \frac{(0.500 \text{ kg})(900 \text{ J/kg}^{\circ}\text{C})(150^{\circ}\text{C}) + (0.250 \text{ kg})(4186 \text{ J/kg}^{\circ}\text{C})(20.0^{\circ}\text{C})}{(0.500 \text{ kg})(900 \text{ J/kg}^{\circ}\text{C}) + (0.250 \text{ kg})(4186 \text{ J/kg}^{\circ}\text{C})} \\&= \frac{88430 \text{ J}}{1496.5 \text{ J}^{\circ}\text{C}} \\&= 59.1^{\circ}\text{C}.\end{aligned}$$

Discussion

This is a typical *calorimetry* problem—two bodies at different temperatures are brought in contact with each other and exchange heat until a common temperature is reached. Why is the final temperature so much closer to 20.0°C than 150°C ? The reason is that water has a greater specific heat than most common substances and thus undergoes a small temperature change for a given heat transfer. A large body of water, such as a lake, requires a large amount of heat to increase its temperature appreciably. This explains why the temperature of a lake stays relatively constant during a day even when the temperature change of the air is large. However, the water temperature does change over longer times (e.g., summer to winter).

Note:**Take-Home Experiment: Temperature Change of Land and Water**

What heats faster, land or water?

To study differences in heat capacity:

- Place equal masses of dry sand (or soil) and water at the same temperature into two small jars. (The average density of soil or sand is about 1.6 times that of water, so you can achieve approximately equal masses by using 50% more water by volume.)
- Heat both (using an oven or a heat lamp) for the same amount of time.
- Record the final temperature of the two masses.
- Now bring both jars to the same temperature by heating for a longer period of time.
- Remove the jars from the heat source and measure their temperature every 5 minutes for about 30 minutes.

Which sample cools off the fastest? This activity replicates the phenomena responsible for land breezes and sea breezes.

Exercise:**Check Your Understanding****Problem:**

If 25 kJ is necessary to raise the temperature of a block from 25°C to 30°C, how much heat is necessary to heat the block from 45°C to 50°C?

Solution:

The heat transfer depends only on the temperature difference. Since the temperature differences are the same in both cases, the same 25 kJ is necessary in the second case.

Summary

- The transfer of heat Q that leads to a change ΔT in the temperature of a body with mass m is $Q = mc\Delta T$, where c is the specific heat of the material. This relationship can also be considered as the definition of specific heat.

Conceptual Questions**Exercise:**

Problem:

What three factors affect the heat transfer that is necessary to change an object's temperature?

Exercise:**Problem:**

The brakes in a car increase in temperature by ΔT when bringing the car to rest from a speed v . How much greater would ΔT be if the car initially had twice the speed? You may assume the car to stop sufficiently fast so that no heat transfers out of the brakes.

Problems & Exercises**Exercise:****Problem:**

On a hot day, the temperature of an 80,000-L swimming pool increases by 1.50°C . What is the net heat transfer during this heating? Ignore any complications, such as loss of water by evaporation.

Solution:**Equation:**

$$5.02 \times 10^8 \text{ J}$$

Exercise:

Problem: Show that $1 \text{ cal/g} \cdot ^\circ\text{C} = 1 \text{ kcal/kg} \cdot ^\circ\text{C}$.

Exercise:**Problem:**

To sterilize a 50.0-g glass baby bottle, we must raise its temperature from 22.0°C to 95.0°C . How much heat transfer is required?

Solution:**Equation:**

$$3.07 \times 10^3 \text{ J}$$

Exercise:**Problem:**

The same heat transfer into identical masses of different substances produces different temperature changes. Calculate the final temperature when 1.00 kcal of heat transfers into 1.00 kg of the following, originally at 20.0°C: (a) water; (b) concrete; (c) steel; and (d) mercury.

Exercise:**Problem:**

Rubbing your hands together warms them by converting work into thermal energy. If a woman rubs her hands back and forth for a total of 20 rubs, at a distance of 7.50 cm per rub, and with an average frictional force of 40.0 N, what is the temperature increase? The mass of tissues warmed is only 0.100 kg, mostly in the palms and fingers.

Solution:**Equation:**

$$0.171^{\circ}\text{C}$$

Exercise:**Problem:**

A 0.250-kg block of a pure material is heated from 20.0°C to 65.0°C by the addition of 4.35 kJ of energy. Calculate its specific heat and identify the substance of which it is most likely composed.

Exercise:**Problem:**

Suppose identical amounts of heat transfer into different masses of copper and water, causing identical changes in temperature. What is the ratio of the mass of copper to water?

Solution:

$$10.8$$

Exercise:

Problem:

(a) The number of kilocalories in food is determined by calorimetry techniques in which the food is burned and the amount of heat transfer is measured. How many kilocalories per gram are there in a 5.00-g peanut if the energy from burning it is transferred to 0.500 kg of water held in a 0.100-kg aluminum cup, causing a 54.9°C temperature increase? (b) Compare your answer to labeling information found on a package of peanuts and comment on whether the values are consistent.

Exercise:**Problem:**

Following vigorous exercise, the body temperature of an 80.0-kg person is 40.0°C . At what rate in watts must the person transfer thermal energy to reduce the body temperature to 37.0°C in 30.0 min, assuming the body continues to produce energy at the rate of 150 W? (1 watt = 1 joule/second or $1\text{ W} = 1\text{ J/s}$).

Solution:

617 W

Exercise:**Problem:**

Even when shut down after a period of normal use, a large commercial nuclear reactor transfers thermal energy at the rate of 150 MW by the radioactive decay of fission products. This heat transfer causes a rapid increase in temperature if the cooling system fails

(1 watt = 1 joule/second or $1\text{ W} = 1\text{ J/s}$ and $1\text{ MW} = 1\text{ megawatt}$). (a) Calculate the rate of temperature increase in degrees Celsius per second ($^{\circ}\text{C/s}$) if the mass of the reactor core is $1.60 \times 10^5\text{ kg}$ and it has an average specific heat of $0.3349\text{ kJ/kg}^{\circ}\text{C}$. (b) How long would it take to obtain a temperature increase of 2000°C , which could cause some metals holding the radioactive materials to melt? (The initial rate of temperature increase would be greater than that calculated here because the heat transfer is concentrated in a smaller mass. Later, however, the temperature increase would slow down because the $5 \times 10^5\text{-kg}$ steel containment vessel would also begin to heat up.)



Radioactive spent-fuel pool at a nuclear power plant. Spent fuel stays hot for a long time. (credit: U.S. Department of Energy)

Glossary

specific heat

the amount of heat necessary to change the temperature of 1.00 kg of a substance by 1.00 °C

Phase Change and Latent Heat

- Examine heat transfer.
- Calculate final temperature from heat transfer.

So far we have discussed temperature change due to heat transfer. No temperature change occurs from heat transfer if ice melts and becomes liquid water (i.e., during a phase change). For example, consider water dripping from icicles melting on a roof warmed by the Sun. Conversely, water freezes in an ice tray cooled by lower-temperature surroundings.



Heat from the air transfers to the ice causing it to melt.
(credit: Mike Brand)

Energy is required to melt a solid because the cohesive bonds between the molecules in the solid must be broken apart such that, in the liquid, the molecules can move around at comparable kinetic energies; thus, there is no rise in temperature. Similarly, energy is needed to vaporize a liquid, because molecules in a liquid interact with each other via attractive forces. There is no temperature change until a phase change is complete. The temperature of a cup of soda initially at 0°C stays at 0°C until all the ice has melted. Conversely, energy is released during freezing and condensation, usually in the form of thermal energy. Work is done by cohesive forces when molecules are brought together. The corresponding energy must be given off (dissipated) to allow them to stay together [\[link\]](#).

The energy involved in a phase change depends on two major factors: the number and strength of bonds or force pairs. The number of bonds is proportional to the number of molecules and thus to the mass of the sample. The strength of forces depends on the type of molecules. The heat Q required to change the phase of a sample of mass m is given by

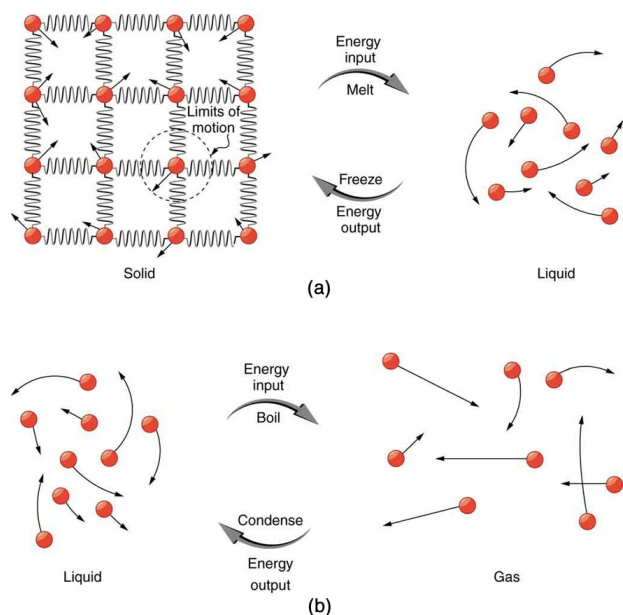
Equation:

$$Q = mL_f \text{ (melting/freezing),}$$

Equation:

$$Q = mL_v \text{ (vaporization/condensation),}$$

where the latent heat of fusion, L_f , and latent heat of vaporization, L_v , are material constants that are determined experimentally. See ([link](#)).



(a) Energy is required to partially overcome the attractive forces between molecules in a solid to form a liquid. That same energy must be removed for freezing to take place. (b) Molecules are separated by large distances when going from liquid to vapor, requiring significant energy to overcome molecular attraction. The same energy must be removed for condensation to take place. There is no temperature change until a phase change is complete.

Latent heat is measured in units of J/kg. Both L_f and L_v depend on the substance, particularly on the strength of its molecular forces as noted earlier. L_f and L_v are collectively called **latent heat coefficients**. They are *latent*, or hidden, because in phase changes, energy enters or leaves a system without causing a temperature change in the system; so, in effect, the energy is hidden. [\[link\]](#) lists representative values of L_f and L_v , together with melting and boiling points.

The table shows that significant amounts of energy are involved in phase changes. Let us look, for example, at how much energy is needed to melt a kilogram of ice at 0°C to produce a kilogram of water at 0°C . Using the equation for a change in temperature and the value for water from [\[link\]](#), we find that $Q = mL_f = (1.0\text{ kg})(334\text{ kJ/kg}) = 334\text{ kJ}$ is the energy to melt a kilogram of ice. This is a lot of energy as it represents the same amount of energy needed to raise the temperature of 1 kg of liquid water from 0°C to 79.8°C . Even more energy is required to vaporize water; it would take 2256 kJ to change 1 kg of liquid water at the normal boiling point (100°C at atmospheric pressure) to steam (water vapor). This example shows that the energy for a phase change is enormous compared to energy associated with temperature changes without a phase change.

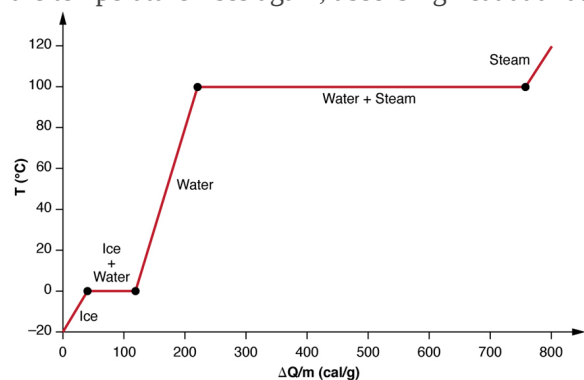
		L_f			L_v	
Substance	Melting point (°C)	kJ/kg	kcal/kg	Boiling point (°C)	kJ/kg	kcal/kg
Helium	−269.7	5.23	1.25	−268.9	20.9	4.99
Hydrogen	−259.3	58.6	14.0	−252.9	452	108
Nitrogen	−210.0	25.5	6.09	−195.8	201	48.0
Oxygen	−218.8	13.8	3.30	−183.0	213	50.9
Ethanol	−114	104	24.9	78.3	854	204
Ammonia	−75		108	−33.4	1370	327
Mercury	−38.9	11.8	2.82	357	272	65.0
Water	0.00	334	79.8	100.0	2256 ^[footnote] At 37.0°C (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg	539 ^[footnote] At 37.0°C (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg
Sulfur	119	38.1	9.10	444.6	326	77.9
Lead	327	24.5	5.85	1750	871	208
Antimony	631	165	39.4	1440	561	134
Aluminum	660	380	90	2450	11400	2720
Silver	961	88.3	21.1	2193	2336	558
Gold	1063	64.5	15.4	2660	1578	377
Copper	1083	134	32.0	2595	5069	1211
Uranium	1133	84	20	3900	1900	454
Tungsten	3410	184	44	5900	4810	1150

Heats of Fusion and Vaporization ^[footnote]

Values quoted at the normal melting and boiling temperatures at standard atmospheric pressure (1 atm).

Phase changes can have a tremendous stabilizing effect even on temperatures that are not near the melting and boiling points, because evaporation and condensation (conversion of a gas into a liquid state) occur even at temperatures below the boiling point. Take, for example, the fact that air temperatures in humid climates rarely go above 35.0°C , which is because most heat transfer goes into evaporating water into the air. Similarly, temperatures in humid weather rarely fall below the dew point because enormous heat is released when water vapor condenses.

We examine the effects of phase change more precisely by considering adding heat into a sample of ice at -20°C ([link](#)). The temperature of the ice rises linearly, absorbing heat at a constant rate of $0.50\text{ cal/g} \cdot ^{\circ}\text{C}$ until it reaches 0°C . Once at this temperature, the ice begins to melt until all the ice has melted, absorbing 79.8 cal/g of heat. The temperature remains constant at 0°C during this phase change. Once all the ice has melted, the temperature of the liquid water rises, absorbing heat at a new constant rate of $1.00\text{ cal/g} \cdot ^{\circ}\text{C}$. At 100°C , the water begins to boil and the temperature again remains constant while the water absorbs 539 cal/g of heat during this phase change. When all the liquid has become steam vapor, the temperature rises again, absorbing heat at a rate of $0.482\text{ cal/g} \cdot ^{\circ}\text{C}$.



A graph of temperature versus energy added. The system is constructed so that no vapor evaporates while ice warms to become liquid water, and so that, when vaporization occurs, the vapor remains in of the system. The long stretches of constant temperature values at 0°C and 100°C reflect the large latent heat of melting and vaporization, respectively.

Water can evaporate at temperatures below the boiling point. More energy is required than at the boiling point, because the kinetic energy of water molecules at temperatures below 100°C is less than that at 100°C , hence less energy is available from random thermal motions. Take, for example, the fact that, at body temperature, perspiration from the skin requires a heat input of 2428 kJ/kg , which is about 10 percent higher than the latent heat of vaporization at 100°C . This heat comes from the skin, and thus provides an effective cooling mechanism in hot weather. High humidity inhibits evaporation, so that body temperature might rise, leaving unevaporated sweat on your brow.

Example:
Calculate Final Temperature from Phase Change: Cooling Soda with Ice Cubes

Three ice cubes are used to chill a soda at 20°C with mass $m_{\text{soda}} = 0.25 \text{ kg}$. The ice is at 0°C and each ice cube has a mass of 6.0 g. Assume that the soda is kept in a foam container so that heat loss can be ignored. Assume the soda has the same heat capacity as water. Find the final temperature when all ice has melted.

Strategy

The ice cubes are at the melting temperature of 0°C. Heat is transferred from the soda to the ice for melting. Melting of ice occurs in two steps: first the phase change occurs and solid (ice) transforms into liquid water at the melting temperature, then the temperature of this water rises. Melting yields water at 0°C, so more heat is transferred from the soda to this water until the water plus soda system reaches thermal equilibrium,

Equation:

$$Q_{\text{ice}} = -Q_{\text{soda}}.$$

The heat transferred to the ice is $Q_{\text{ice}} = m_{\text{ice}}L_f + m_{\text{ice}}c_W(T_f - 0^\circ\text{C})$. The heat given off by the soda is $Q_{\text{soda}} = m_{\text{soda}}c_W(T_f - 20^\circ\text{C})$. Since no heat is lost, $Q_{\text{ice}} = -Q_{\text{soda}}$, so that

Equation:

$$m_{\text{ice}}L_f + m_{\text{ice}}c_W(T_f - 0^\circ\text{C}) = -m_{\text{soda}}c_W(T_f - 20^\circ\text{C}).$$

Bring all terms involving T_f on the left-hand-side and all other terms on the right-hand-side. Solve for the unknown quantity T_f :

Equation:

$$T_f = \frac{m_{\text{soda}}c_W(20^\circ\text{C}) - m_{\text{ice}}L_f}{(m_{\text{soda}} + m_{\text{ice}})c_W}.$$

Solution

1. Identify the known quantities. The mass of ice is $m_{\text{ice}} = 3 \times 6.0 \text{ g} = 0.018 \text{ kg}$ and the mass of soda is $m_{\text{soda}} = 0.25 \text{ kg}$.

2. Calculate the terms in the numerator:

Equation:

$$m_{\text{soda}}c_W(20^\circ\text{C}) = (0.25 \text{ kg})(4186 \text{ J/kg} \cdot ^\circ\text{C})(20^\circ\text{C}) = 20,930 \text{ J}$$

and

Equation:

$$m_{\text{ice}}L_f = (0.018 \text{ kg})(334,000 \text{ J/kg}) = 6012 \text{ J}.$$

3. Calculate the denominator:

Equation:

$$(m_{\text{soda}} + m_{\text{ice}})c_W = (0.25 \text{ kg} + 0.018 \text{ kg})(4186 \text{ J/(kg} \cdot ^\circ\text{C)}) = 1122 \text{ J/}^\circ\text{C}.$$

4. Calculate the final temperature:

Equation:

$$T_f = \frac{20,930 \text{ J} - 6012 \text{ J}}{1122 \text{ J/}^\circ\text{C}} = 13^\circ\text{C}.$$

Discussion

This example illustrates the enormous energies involved during a phase change. The mass of ice is about 7 percent the mass of water but leads to a noticeable change in the temperature of soda. Although we assumed that the ice was at the freezing temperature, this is incorrect: the typical temperature is -6°C . However, this correction gives a final temperature that is essentially identical to the result we found. Can you explain why?

We have seen that vaporization requires heat transfer to a liquid from the surroundings, so that energy is released by the surroundings. Condensation is the reverse process, increasing the temperature of the surroundings. This increase may seem surprising, since we associate condensation with cold objects—the glass in the figure, for example. However, energy must be removed from the condensing molecules to make a vapor condense. The energy is exactly the same as that required to make the phase change in the other direction, from liquid to vapor, and so it can be calculated from $Q = mL_v$.



Condensation forms on this glass of iced tea because the temperature of the nearby air is reduced to below the dew point. The rate at which water molecules join together exceeds the rate at which they separate, and so water condenses. Energy is released when the water condenses, speeding the melting of the ice in the glass. (credit: Jenny Downing)

Note:

Real-World Application

Energy is also released when a liquid freezes. This phenomenon is used by fruit growers in Florida to protect oranges when the temperature is close to the freezing point (0°C). Growers spray water on the

plants in orchards so that the water freezes and heat is released to the growing oranges on the trees. This prevents the temperature inside the orange from dropping below freezing, which would damage the fruit.



The ice on these trees released large amounts of energy when it froze, helping to prevent the temperature of the trees from dropping below 0°C .

Water is intentionally sprayed on orchards to help prevent hard frosts.

(credit: Hermann Hammer)

Sublimation is the transition from solid to vapor phase. You may have noticed that snow can disappear into thin air without a trace of liquid water, or the disappearance of ice cubes in a freezer. The reverse is also true: Frost can form on very cold windows without going through the liquid stage. A popular effect is the making of “smoke” from dry ice, which is solid carbon dioxide. Sublimation occurs because the equilibrium vapor pressure of solids is not zero. Certain air fresheners use the sublimation of a solid to inject a perfume into the room. Moth balls are a slightly toxic example of a phenol (an organic compound) that sublimates, while some solids, such as osmium tetroxide, are so toxic that they must be kept in sealed containers to prevent human exposure to their sublimation-produced vapors.



(a)



(b)

Direct transitions
between solid and

vapor are common,
sometimes useful,
and even beautiful.

(a) Dry ice
sublimates directly to
carbon dioxide gas.
The visible vapor is
made of water
droplets. (credit:
Windell Oskay) (b)
Frost forms patterns
on a very cold
window, an example
of a solid formed
directly from a
vapor. (credit: Liz
West)

All phase transitions involve heat. In the case of direct solid-vapor transitions, the energy required is given by the equation $Q = mL_s$, where L_s is the **heat of sublimation**, which is the energy required to change 1.00 kg of a substance from the solid phase to the vapor phase. L_s is analogous to L_f and L_v , and its value depends on the substance. Sublimation requires energy input, so that dry ice is an effective coolant, whereas the reverse process (i.e., frosting) releases energy. The amount of energy required for sublimation is of the same order of magnitude as that for other phase transitions.

The material presented in this section and the preceding section allows us to calculate any number of effects related to temperature and phase change. In each case, it is necessary to identify which temperature and phase changes are taking place and then to apply the appropriate equation. Keep in mind that heat transfer and work can cause both temperature and phase changes.

Problem-Solving Strategies for the Effects of Heat Transfer

1. *Examine the situation to determine that there is a change in the temperature or phase. Is there heat transfer into or out of the system?* When the presence or absence of a phase change is not obvious, you may wish to first solve the problem as if there were no phase changes, and examine the temperature change obtained. If it is sufficient to take you past a boiling or melting point, you should then go back and do the problem in steps—temperature change, phase change, subsequent temperature change, and so on.
2. *Identify and list all objects that change temperature and phase.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns).* A written list is useful.
4. *Make a list of what is given or what can be inferred from the problem as stated (identify the knowns).*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).* If there is a temperature change, the transferred heat depends on the specific heat (see [\[link\]](#)) whereas, for a phase change, the transferred heat depends on the latent heat. See [\[link\]](#).
6. *Substitute the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.* You will need to do this in steps if there is more than one stage to the process (such as a temperature change followed by a phase change).

7. Check the answer to see if it is reasonable: Does it make sense? As an example, be certain that the temperature change does not also cause a phase change that you have not taken into account.

Exercise:

Check Your Understanding

Problem:

Why does snow remain on mountain slopes even when daytime temperatures are higher than the freezing temperature?

Solution:

Snow is formed from ice crystals and thus is the solid phase of water. Because enormous heat is necessary for phase changes, it takes a certain amount of time for this heat to be accumulated from the air, even if the air is above 0°C. The warmer the air is, the faster this heat exchange occurs and the faster the snow melts.

Summary

- Most substances can exist either in solid, liquid, and gas forms, which are referred to as “phases.”
- Phase changes occur at fixed temperatures for a given substance at a given pressure, and these temperatures are called boiling and freezing (or melting) points.
- During phase changes, heat absorbed or released is given by:

Equation:

$$Q = mL,$$

where L is the latent heat coefficient.

Conceptual Questions

Exercise:

Problem:

Heat transfer can cause temperature and phase changes. What else can cause these changes?

Exercise:

Problem:

How does the latent heat of fusion of water help slow the decrease of air temperatures, perhaps preventing temperatures from falling significantly below 0°C, in the vicinity of large bodies of water?

Exercise:

Problem: What is the temperature of ice right after it is formed by freezing water?

Exercise:

Problem:

If you place 0°C ice into 0°C water in an insulated container, what will happen? Will some ice melt, will more water freeze, or will neither take place?

Exercise:

Problem:

What effect does condensation on a glass of ice water have on the rate at which the ice melts? Will the condensation speed up the melting process or slow it down?

Exercise:

Problem:

In very humid climates where there are numerous bodies of water, such as in Florida, it is unusual for temperatures to rise above about 35°C (95°F). In deserts, however, temperatures can rise far above this. Explain how the evaporation of water helps limit high temperatures in humid climates.

Exercise:

Problem:

In winters, it is often warmer in San Francisco than in nearby Sacramento, 150 km inland. In summers, it is nearly always hotter in Sacramento. Explain how the bodies of water surrounding San Francisco moderate its extreme temperatures.

Exercise:

Problem:

Putting a lid on a boiling pot greatly reduces the heat transfer necessary to keep it boiling. Explain why.

Exercise:

Problem:

Freeze-dried foods have been dehydrated in a vacuum. During the process, the food freezes and must be heated to facilitate dehydration. Explain both how the vacuum speeds up dehydration and why the food freezes as a result.

Exercise:

Problem:

When still air cools by radiating at night, it is unusual for temperatures to fall below the dew point. Explain why.

Exercise:

Problem:

In a physics classroom demonstration, an instructor inflates a balloon by mouth and then cools it in liquid nitrogen. When cold, the shrunken balloon has a small amount of light blue liquid in it, as well as some snow-like crystals. As it warms up, the liquid boils, and part of the crystals sublime, with some crystals lingering for awhile and then producing a liquid. Identify the blue liquid and the two solids in the cold balloon. Justify your identifications using data from [\[link\]](#).

Problems & Exercises

Exercise:

Problem:

How much heat transfer (in kilocalories) is required to thaw a 0.450-kg package of frozen vegetables originally at 0°C if their heat of fusion is the same as that of water?

Solution:

35.9 kcal

Exercise:

Problem:

A bag containing 0°C ice is much more effective in absorbing energy than one containing the same amount of 0°C water.

- How much heat transfer is necessary to raise the temperature of 0.800 kg of water from 0°C to 30.0°C?
- How much heat transfer is required to first melt 0.800 kg of 0°C ice and then raise its temperature?
- Explain how your answer supports the contention that the ice is more effective.

Exercise:

Problem:

- How much heat transfer is required to raise the temperature of a 0.750-kg aluminum pot containing 2.50 kg of water from 30.0°C to the boiling point and then boil away 0.750 kg of water?
 - How long does this take if the rate of heat transfer is 500 W
1 watt = 1 joule/second (1 W = 1 J/s)?
-

Solution:

(a) 591 kcal

(b) 4.94×10^3 s

Exercise:

Problem:

The formation of condensation on a glass of ice water causes the ice to melt faster than it would otherwise. If 8.00 g of condensation forms on a glass containing both water and 200 g of ice, how many grams of the ice will melt as a result? Assume no other heat transfer occurs.

Exercise:

Problem:

On a trip, you notice that a 3.50-kg bag of ice lasts an average of one day in your cooler. What is the average power in watts entering the ice if it starts at 0°C and completely melts to 0°C water in exactly one day 1 watt = 1 joule/second (1 W = 1 J/s)?

Solution:

13.5 W

Exercise:**Problem:**

On a certain dry sunny day, a swimming pool's temperature would rise by 1.50°C if not for evaporation. What fraction of the water must evaporate to carry away precisely enough energy to keep the temperature constant?

Exercise:**Problem:**

- (a) How much heat transfer is necessary to raise the temperature of a 0.200-kg piece of ice from -20.0°C to 130°C , including the energy needed for phase changes?
- (b) How much time is required for each stage, assuming a constant 20.0 kJ/s rate of heat transfer?
- (c) Make a graph of temperature versus time for this process.

Solution:

(a) 148 kcal

(b) 0.418 s, 3.34 s, 4.19 s, 22.6 s, 0.456 s

Exercise:**Problem:**

In 1986, a gargantuan iceberg broke away from the Ross Ice Shelf in Antarctica. It was approximately a rectangle 160 km long, 40.0 km wide, and 250 m thick.

- (a) What is the mass of this iceberg, given that the density of ice is 917 kg/m^3 ?
- (b) How much heat transfer (in joules) is needed to melt it?
- (c) How many years would it take sunlight alone to melt ice this thick, if the ice absorbs an average of 100 W/m^2 , 12.00 h per day?

Exercise:**Problem:**

How many grams of coffee must evaporate from 350 g of coffee in a 100-g glass cup to cool the coffee from 95.0°C to 45.0°C ? You may assume the coffee has the same thermal properties as water and that the average heat of vaporization is 2340 kJ/kg (560 cal/g). (You may neglect the change in mass of the coffee as it cools, which will give you an answer that is slightly larger than correct.)

Solution:

33.0 g

Exercise:

Problem:

- (a) It is difficult to extinguish a fire on a crude oil tanker, because each liter of crude oil releases 2.80×10^7 J of energy when burned. To illustrate this difficulty, calculate the number of liters of water that must be expended to absorb the energy released by burning 1.00 L of crude oil, if the water has its temperature raised from 20.0°C to 100°C , it boils, and the resulting steam is raised to 300°C .
- (b) Discuss additional complications caused by the fact that crude oil has a smaller density than water.
-

Solution:

(a) 9.67 L

(b) Crude oil is less dense than water, so it floats on top of the water, thereby exposing it to the oxygen in the air, which it uses to burn. Also, if the water is under the oil, it is less efficient in absorbing the heat generated by the oil.

Exercise:**Problem:**

The energy released from condensation in thunderstorms can be very large. Calculate the energy released into the atmosphere for a small storm of radius 1 km, assuming that 1.0 cm of rain is precipitated uniformly over this area.

Exercise:

Problem: To help prevent frost damage, 4.00 kg of 0°C water is sprayed onto a fruit tree.

- (a) How much heat transfer occurs as the water freezes?
- (b) How much would the temperature of the 200-kg tree decrease if this amount of heat transferred from the tree? Take the specific heat to be $3.35 \text{ kJ/kg} \cdot ^\circ\text{C}$, and assume that no phase change occurs.
-

Solution:

a) 319 kcal

b) 2.00°C

Exercise:**Problem:**

A 0.250-kg aluminum bowl holding 0.800 kg of soup at 25.0°C is placed in a freezer. What is the final temperature if 377 kJ of energy is transferred from the bowl and soup, assuming the soup's thermal properties are the same as that of water? Explicitly show how you follow the steps in [Problem-Solving Strategies for the Effects of Heat Transfer](#).

Exercise:**Problem:**

A 0.0500-kg ice cube at -30.0°C is placed in 0.400 kg of 35.0°C water in a very well-insulated container. What is the final temperature?

Solution:

20.6°C

Exercise:**Problem:**

If you pour 0.0100 kg of 20.0°C water onto a 1.20-kg block of ice (which is initially at -15.0°C), what is the final temperature? You may assume that the water cools so rapidly that effects of the surroundings are negligible.

Exercise:**Problem:**

Indigenous people sometimes cook in watertight baskets by placing hot rocks into water to bring it to a boil. What mass of 500°C rock must be placed in 4.00 kg of 15.0°C water to bring its temperature to 100°C, if 0.0250 kg of water escapes as vapor from the initial sizzle? You may neglect the effects of the surroundings and take the average specific heat of the rocks to be that of granite.

Solution:

4.38 kg

Exercise:**Problem:**

What would be the final temperature of the pan and water in [Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan](#) if 0.260 kg of water was placed in the pan and 0.0100 kg of the water evaporated immediately, leaving the remainder to come to a common temperature with the pan?

Exercise:**Problem:**

In some countries, liquid nitrogen is used on dairy trucks instead of mechanical refrigerators. A 3.00-hour delivery trip requires 200 L of liquid nitrogen, which has a density of 808 kg/m^3 .

(a) Calculate the heat transfer necessary to evaporate this amount of liquid nitrogen and raise its temperature to 3.00°C . (Use c_p and assume it is constant over the temperature range.) This value is the amount of cooling the liquid nitrogen supplies.

(b) What is this heat transfer rate in kilowatt-hours?

(c) Compare the amount of cooling obtained from melting an identical mass of 0°C ice with that from evaporating the liquid nitrogen.

Solution:

(a) $1.57 \times 10^4\text{ kcal}$

(b) $18.3\text{ kW} \cdot \text{h}$

(c) $1.29 \times 10^4\text{ kcal}$

Exercise:**Problem:**

Some gun fanciers make their own bullets, which involves melting and casting the lead slugs. How much heat transfer is needed to raise the temperature and melt 0.500 kg of lead, starting from 25.0°C ?

Glossary

heat of sublimation

the energy required to change a substance from the solid phase to the vapor phase

latent heat coefficient

a physical constant equal to the amount of heat transferred for every 1 kg of a substance during the change in phase of the substance

sublimation

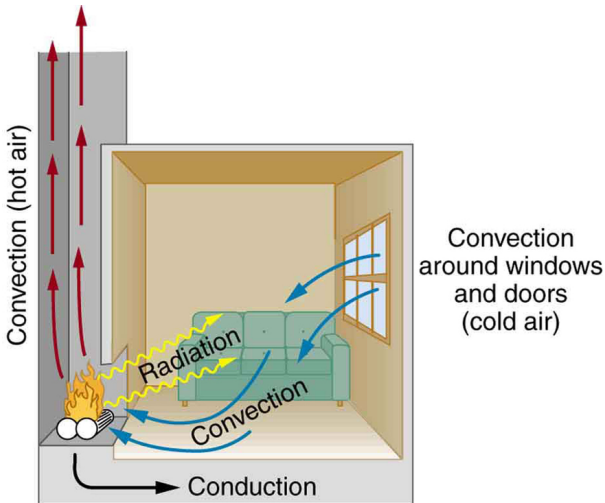
the transition from the solid phase to the vapor phase

Heat Transfer Methods

- Discuss the different methods of heat transfer.

Equally as interesting as the effects of heat transfer on a system are the methods by which this occurs. Whenever there is a temperature difference, heat transfer occurs. Heat transfer may occur rapidly, such as through a cooking pan, or slowly, such as through the walls of a picnic ice chest. We can control rates of heat transfer by choosing materials (such as thick wool clothing for the winter), controlling air movement (such as the use of weather stripping around doors), or by choice of color (such as a white roof to reflect summer sunlight). So many processes involve heat transfer, so that it is hard to imagine a situation where no heat transfer occurs. Yet every process involving heat transfer takes place by only three methods:

1. **Conduction** is heat transfer through stationary matter by physical contact. (The matter is stationary on a macroscopic scale—we know there is thermal motion of the atoms and molecules at any temperature above absolute zero.) Heat transferred between the electric burner of a stove and the bottom of a pan is transferred by conduction.
2. **Convection** is the heat transfer by the macroscopic movement of a fluid. This type of transfer takes place in a forced-air furnace and in weather systems, for example.
3. Heat transfer by **radiation** occurs when microwaves, infrared radiation, visible light, or another form of electromagnetic radiation is emitted or absorbed. An obvious example is the warming of the Earth by the Sun. A less obvious example is thermal radiation from the human body.



In a fireplace, heat transfer occurs by all three methods: conduction, convection, and radiation. Radiation is responsible for most of the heat transferred into the room. Heat transfer also occurs through conduction into the room, but at a much slower rate. Heat transfer by convection also occurs through cold air entering the room around windows and hot air leaving the room by rising up the chimney.

We examine these methods in some detail in the three following modules. Each method has unique and interesting characteristics, but all three do have one thing in common: they transfer heat solely because of a temperature difference [\[link\]](#).

Exercise:

Check Your Understanding

Problem:

Name an example from daily life (different from the text) for each mechanism of heat transfer.

Solution:

Conduction: Heat transfers into your hands as you hold a hot cup of coffee.

Convection: Heat transfers as the barista “steams” cold milk to make hot *cocoa*.

Radiation: Reheating a cold cup of coffee in a microwave oven.

Summary

- Heat is transferred by three different methods: conduction, convection, and radiation.

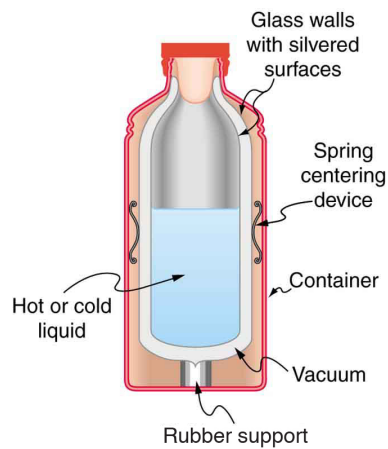
Conceptual Questions**Exercise:****Problem:**

What are the main methods of heat transfer from the hot core of Earth to its surface? From Earth’s surface to outer space?

When our bodies get too warm, they respond by sweating and increasing blood circulation to the surface to transfer thermal energy away from the core. What effect will this have on a person in a ° hot tub?

[\[link\]](#) shows a cut-away drawing of a thermos bottle (also known as a Dewar flask), which is a device designed specifically to slow down all forms of heat transfer. Explain the functions of the various parts, such as the

vacuum, the silvering of the walls, the thin-walled long glass neck, the rubber support, the air layer, and the stopper.



The construction of a thermos bottle is designed to inhibit all methods of heat transfer.

Glossary

conduction

heat transfer through stationary matter by physical contact

convection

heat transfer by the macroscopic movement of fluid

radiation

heat transfer which occurs when microwaves, infrared radiation, visible light, or other electromagnetic radiation is emitted or absorbed

Conduction

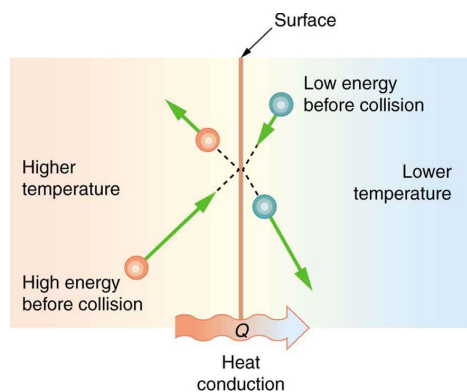
- Calculate thermal conductivity.
- Observe conduction of heat in collisions.
- Study thermal conductivities of common substances.



Insulation is used to limit the conduction of heat from the inside to the outside (in winters) and from the outside to the inside (in summers).
(credit: Giles Douglas)

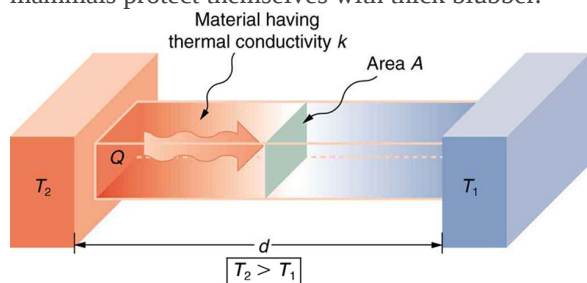
Your feet feel cold as you walk barefoot across the living room carpet in your cold house and then step onto the kitchen tile floor. This result is intriguing, since the carpet and tile floor are both at the same temperature. The different sensation you feel is explained by the different rates of heat transfer: the heat loss during the same time interval is greater for skin in contact with the tiles than with the carpet, so the temperature drop is greater on the tiles.

Some materials conduct thermal energy faster than others. In general, good conductors of electricity (metals like copper, aluminum, gold, and silver) are also good heat conductors, whereas insulators of electricity (wood, plastic, and rubber) are poor heat conductors. [\[link\]](#) shows molecules in two bodies at different temperatures. The (average) kinetic energy of a molecule in the hot body is higher than in the colder body. If two molecules collide, an energy transfer from the molecule with greater kinetic energy to the molecule with less kinetic energy occurs. The cumulative effect from all collisions results in a net flux of heat from the hot body to the colder body. The heat flux thus depends on the temperature difference $\Delta T = T_{\text{hot}} - T_{\text{cold}}$. Therefore, you will get a more severe burn from boiling water than from hot tap water. Conversely, if the temperatures are the same, the net heat transfer rate falls to zero, and equilibrium is achieved. Owing to the fact that the number of collisions increases with increasing area, heat conduction depends on the cross-sectional area. If you touch a cold wall with your palm, your hand cools faster than if you just touch it with your fingertip.



The molecules in two bodies at different temperatures have different average kinetic energies. Collisions occurring at the contact surface tend to transfer energy from high-temperature regions to low-temperature regions. In this illustration, a molecule in the lower temperature region (right side) has low energy before collision, but its energy increases after colliding with the contact surface. In contrast, a molecule in the higher temperature region (left side) has high energy before collision, but its energy decreases after colliding with the contact surface.

A third factor in the mechanism of conduction is the thickness of the material through which heat transfers. The figure below shows a slab of material with different temperatures on either side. Suppose that T_2 is greater than T_1 , so that heat is transferred from left to right. Heat transfer from the left side to the right side is accomplished by a series of molecular collisions. The thicker the material, the more time it takes to transfer the same amount of heat. This model explains why thick clothing is warmer than thin clothing in winters, and why Arctic mammals protect themselves with thick blubber.



Heat conduction occurs through any material, represented here by a rectangular bar, whether window glass or walrus blubber. The temperature of the material is T_2 on the left and T_1 on the right, where T_2 is greater than T_1 .

The rate of heat transfer by conduction is directly proportional to the surface area A , the temperature difference $T_2 - T_1$, and the substance's conductivity k . The rate of heat transfer is inversely proportional to the thickness d .

Lastly, the heat transfer rate depends on the material properties described by the coefficient of thermal conductivity. All four factors are included in a simple equation that was deduced from and is confirmed by experiments. The **rate of conductive heat transfer** through a slab of material, such as the one in [\[link\]](#), is given by

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d},$$

where Q/t is the rate of heat transfer in watts or kilocalories per second, k is the **thermal conductivity** of the material, A and d are its surface area and thickness, as shown in [\[link\]](#), and $(T_2 - T_1)$ is the temperature difference across the slab. [\[link\]](#) gives representative values of thermal conductivity.

Example:

Calculating Heat Transfer Through Conduction: Conduction Rate Through an Ice Box

A Styrofoam ice box has a total area of 0.950 m^2 and walls with an average thickness of 2.50 cm . The box contains ice, water, and canned beverages at 0°C . The inside of the box is kept cold by melting ice. How much ice melts in one day if the ice box is kept in the trunk of a car at 35.0°C ?

Strategy

This question involves both heat for a phase change (melting of ice) and the transfer of heat by conduction. To find the amount of ice melted, we must find the net heat transferred. This value can be obtained by calculating the rate of heat transfer by conduction and multiplying by time.

Solution

1. Identify the knowns.

Equation:

$$A = 0.950 \text{ m}^2; d = 2.50 \text{ cm} = 0.0250 \text{ m}; T_1 = 0^\circ\text{C}; T_2 = 35.0^\circ\text{C}; t = 1 \text{ day} = 24 \text{ hours} = 86,400 \text{ s}.$$

2. Identify the unknowns. We need to solve for the mass of the ice, m . We will also need to solve for the net heat transferred to melt the ice, Q .
3. Determine which equations to use. The rate of heat transfer by conduction is given by

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}.$$

4. The heat is used to melt the ice: $Q = mL_f$.
5. Insert the known values:

Equation:

$$\frac{Q}{t} = \frac{(0.010 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C})(0.950 \text{ m}^2)(35.0^\circ\text{C} - 0^\circ\text{C})}{0.0250 \text{ m}} = 13.3 \text{ J/s}.$$

6. Multiply the rate of heat transfer by the time (1 day = 86,400 s):

Equation:

$$Q = (Q/t)t = (13.3 \text{ J/s})(86,400 \text{ s}) = 1.15 \times 10^6 \text{ J}.$$

7. Set this equal to the heat transferred to melt the ice: $Q = mL_f$. Solve for the mass m :

Equation:

$$m = \frac{Q}{L_f} = \frac{1.15 \times 10^6 \text{ J}}{334 \times 10^3 \text{ J/kg}} = 3.44 \text{ kg}.$$

Discussion

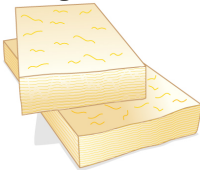
The result of 3.44 kg, or about 7.6 lbs, seems about right, based on experience. You might expect to use about a 4 kg (7–10 lb) bag of ice per day. A little extra ice is required if you add any warm food or beverages. Inspecting the conductivities in [\[link\]](#) shows that Styrofoam is a very poor conductor and thus a good insulator. Other good insulators include fiberglass, wool, and goose-down feathers. Like Styrofoam, these all incorporate many small pockets of air, taking advantage of air's poor thermal conductivity.

Substance	Thermal conductivity k (J/s·m·°C)
Silver	420
Copper	390
Gold	318
Aluminum	220
Steel iron	80
Steel (stainless)	14
Ice	2.2
Glass (average)	0.84
Concrete brick	0.84
Water	0.6
Fatty tissue (without blood)	0.2
Asbestos	0.16
Plasterboard	0.16
Wood	0.08–0.16

Substance	Thermal conductivity k (J/s·m·°C)
Snow (dry)	0.10
Cork	0.042
Glass wool	0.042
Wool	0.04
Down feathers	0.025
Air	0.023
Styrofoam	0.010

Thermal Conductivities of Common Substances[\[footnote\]](#)
At temperatures near 0°C.

A combination of material and thickness is often manipulated to develop good insulators—the smaller the conductivity k and the larger the thickness d , the better. The ratio of d/k will thus be large for a good insulator. The ratio d/k is called the **R factor**. The rate of conductive heat transfer is inversely proportional to R . The larger the value of R , the better the insulation. R factors are most commonly quoted for household insulation, refrigerators, and the like—unfortunately, it is still in non-metric units of $\text{ft}^2 \cdot ^\circ\text{F} \cdot \text{h} / \text{Btu}$, although the unit usually goes unstated (1 British thermal unit [Btu] is the amount of energy needed to change the temperature of 1.0 lb of water by 1.0 °F). A couple of representative values are an R factor of 11 for 3.5-in-thick fiberglass batts (pieces) of insulation and an R factor of 19 for 6.5-in-thick fiberglass batts. Walls are usually insulated with 3.5-in batts, while ceilings are usually insulated with 6.5-in batts. In cold climates, thicker batts may be used in ceilings and walls.



The fiberglass batt is used for insulation of walls and ceilings to prevent heat transfer between the inside of the building and the outside environment.

Note that in [\[link\]](#), the best thermal conductors—silver, copper, gold, and aluminum—are also the best electrical conductors, again related to the density of free electrons in them. Cooking utensils are typically made

from good conductors.

Example:

Calculating the Temperature Difference Maintained by a Heat Transfer: Conduction Through an Aluminum Pan

Water is boiling in an aluminum pan placed on an electrical element on a stovetop. The sauce pan has a bottom that is 0.800 cm thick and 14.0 cm in diameter. The boiling water is evaporating at the rate of 1.00 g/s. What is the temperature difference across (through) the bottom of the pan?

Strategy

Conduction through the aluminum is the primary method of heat transfer here, and so we use the equation for the rate of heat transfer and solve for the temperature difference.

Equation:

$$T_2 - T_1 = \frac{Q}{t} \left(\frac{d}{kA} \right).$$

Solution

1. Identify the knowns and convert them to the SI units.

The thickness of the pan, $d = 0.800 \text{ cm} = 8.0 \times 10^{-3} \text{ m}$, the area of the pan, $A = \pi(0.14/2)^2 \text{ m}^2 = 1.54 \times 10^{-2} \text{ m}^2$, and the thermal conductivity, $k = 220 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C}$.

2. Calculate the necessary heat of vaporization of 1 g of water:

Equation:

$$Q = mL_v = (1.00 \times 10^{-3} \text{ kg})(2256 \times 10^3 \text{ J/kg}) = 2256 \text{ J}.$$

3. Calculate the rate of heat transfer given that 1 g of water melts in one second:

Equation:

$$Q/t = 2256 \text{ J/s or } 2.26 \text{ kW}.$$

4. Insert the knowns into the equation and solve for the temperature difference:

Equation:

$$T_2 - T_1 = \frac{Q}{t} \left(\frac{d}{kA} \right) = (2256 \text{ J/s}) \frac{8.00 \times 10^{-3} \text{ m}}{(220 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C})(1.54 \times 10^{-2} \text{ m}^2)} = 5.33^\circ\text{C}.$$

Discussion

The value for the heat transfer $Q/t = 2.26 \text{ kW}$ or 2256 J/s is typical for an electric stove. This value gives a remarkably small temperature difference between the stove and the pan. Consider that the stove burner is red hot while the inside of the pan is nearly 100°C because of its contact with boiling water. This contact effectively cools the bottom of the pan in spite of its proximity to the very hot stove burner. Aluminum is such a good conductor that it only takes this small temperature difference to produce a heat transfer of 2.26 kW into the pan.

Conduction is caused by the random motion of atoms and molecules. As such, it is an ineffective mechanism for heat transport over macroscopic distances and short time distances. Take, for example, the temperature on the Earth, which would be unbearably cold during the night and extremely hot during the day if heat transport in the atmosphere was to be only through conduction. In another example, car engines would overheat unless there was a more efficient way to remove excess heat from the pistons.

Exercise:
Check Your Understanding

Problem:

How does the rate of heat transfer by conduction change when all spatial dimensions are doubled?

Solution:

Because area is the product of two spatial dimensions, it increases by a factor of four when each dimension is doubled ($A_{\text{final}} = (2d)^2 = 4d^2 = 4A_{\text{initial}}$). The distance, however, simply doubles. Because the temperature difference and the coefficient of thermal conductivity are independent of the spatial dimensions, the rate of heat transfer by conduction increases by a factor of four divided by two, or two:

Equation:

$$\left(\frac{Q}{t}\right)_{\text{final}} = \frac{kA_{\text{final}}(T_2 - T_1)}{d_{\text{final}}} = \frac{k(4A_{\text{initial}})(T_2 - T_1)}{2d_{\text{initial}}} = 2 \frac{kA_{\text{initial}}(T_2 - T_1)}{d_{\text{initial}}} = 2 \left(\frac{Q}{t}\right)_{\text{initial}}.$$

Summary

- Heat conduction is the transfer of heat between two objects in direct contact with each other.
- The rate of heat transfer Q/t (energy per unit time) is proportional to the temperature difference $T_2 - T_1$ and the contact area A and inversely proportional to the distance d between the objects:

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}.$$

Conceptual Questions

Exercise:

Problem:

Some electric stoves have a flat ceramic surface with heating elements hidden beneath. A pot placed over a heating element will be heated, while it is safe to touch the surface only a few centimeters away. Why is ceramic, with a conductivity less than that of a metal but greater than that of a good insulator, an ideal choice for the stove top?

Exercise:

Problem:

Loose-fitting white clothing covering most of the body is ideal for desert dwellers, both in the hot Sun and during cold evenings. Explain how such clothing is advantageous during both day and night.



A jellabiya is worn by many men in Egypt. (credit: Zerida)

Problems & Exercises

Exercise:

Problem:

- (a) Calculate the rate of heat conduction through house walls that are 13.0 cm thick and that have an average thermal conductivity twice that of glass wool. Assume there are no windows or doors. The surface area of the walls is 120 m^2 and their inside surface is at 18.0°C , while their outside surface is at 5.00°C .
- (b) How many 1-kW room heaters would be needed to balance the heat transfer due to conduction?

Solution:

- (a) $1.01 \times 10^3 \text{ W}$
- (b) One

Exercise:

Problem:

The rate of heat conduction out of a window on a winter day is rapid enough to chill the air next to it. To see just how rapidly the windows transfer heat by conduction, calculate the rate of conduction in watts through a 3.00-m^2 window that is 0.635 cm thick (1/4 in) if the temperatures of the inner and outer surfaces are 5.00°C and -10.0°C , respectively. This rapid rate will not be maintained—the inner surface will cool, and even result in frost formation.

Exercise:

Problem:

Calculate the rate of heat conduction out of the human body, assuming that the core internal temperature is 37.0°C , the skin temperature is 34.0°C , the thickness of the tissues between averages 1.00 cm, and the surface area is 1.40 m^2 .

Solution:

84.0 W

Exercise:**Problem:**

Suppose you stand with one foot on ceramic flooring and one foot on a wool carpet, making contact over an area of 80.0 cm^2 with each foot. Both the ceramic and the carpet are 2.00 cm thick and are 10.0°C on their bottom sides. At what rate must heat transfer occur from each foot to keep the top of the ceramic and carpet at 33.0°C ?

Exercise:**Problem:**

A man consumes 3000 kcal of food in one day, converting most of it to maintain body temperature. If he loses half this energy by evaporating water (through breathing and sweating), how many kilograms of water evaporate?

Solution:

2.59 kg

Exercise:**Problem:**

(a) A firewalker runs across a bed of hot coals without sustaining burns. Calculate the heat transferred by conduction into the sole of one foot of a firewalker given that the bottom of the foot is a 3.00-mm-thick callus with a conductivity at the low end of the range for wood and its density is 300 kg/m^3 . The area of contact is 25.0 cm^2 , the temperature of the coals is 700°C , and the time in contact is 1.00 s.

(b) What temperature increase is produced in the 25.0 cm^3 of tissue affected?

(c) What effect do you think this will have on the tissue, keeping in mind that a callus is made of dead cells?

Exercise:**Problem:**

(a) What is the rate of heat conduction through the 3.00-cm-thick fur of a large animal having a 1.40-m^2 surface area? Assume that the animal's skin temperature is 32.0°C , that the air temperature is -5.00°C , and that fur has the same thermal conductivity as air. (b) What food intake will the animal need in one day to replace this heat transfer?

Solution:

(a) 39.7 W

(b) 820 kcal

Exercise:

Problem:

A walrus transfers energy by conduction through its blubber at the rate of 150 W when immersed in -1.00°C water. The walrus's internal core temperature is 37.0°C , and it has a surface area of 2.00 m^2 . What is the average thickness of its blubber, which has the conductivity of fatty tissues without blood?



Walrus on ice. (credit: Captain Budd Christman, NOAA Corps)

Exercise:**Problem:**

Compare the rate of heat conduction through a 13.0-cm-thick wall that has an area of 10.0 m^2 and a thermal conductivity twice that of glass wool with the rate of heat conduction through a window that is 0.750 cm thick and that has an area of 2.00 m^2 , assuming the same temperature difference across each.

Solution:

35 to 1, window to wall

Exercise:**Problem:**

Suppose a person is covered head to foot by wool clothing with average thickness of 2.00 cm and is transferring energy by conduction through the clothing at the rate of 50.0 W. What is the temperature difference across the clothing, given the surface area is 1.40 m^2 ?

Exercise:**Problem:**

Some stove tops are smooth ceramic for easy cleaning. If the ceramic is 0.600 cm thick and heat conduction occurs through the same area and at the same rate as computed in [\[link\]](#), what is the temperature difference across it? Ceramic has the same thermal conductivity as glass and brick.

Solution:

$1.05 \times 10^3\text{ K}$

Exercise:

Problem:

One easy way to reduce heating (and cooling) costs is to add extra insulation in the attic of a house. Suppose the house already had 15 cm of fiberglass insulation in the attic and in all the exterior surfaces. If you added an extra 8.0 cm of fiberglass to the attic, then by what percentage would the heating cost of the house drop? Take the single story house to be of dimensions 10 m by 15 m by 3.0 m. Ignore air infiltration and heat loss through windows and doors.

Exercise:**Problem:**

(a) Calculate the rate of heat conduction through a double-paned window that has a 1.50-m^2 area and is made of two panes of 0.800-cm-thick glass separated by a 1.00-cm air gap. The inside surface temperature is 15.0°C , while that on the outside is -10.0°C . (Hint: There are identical temperature drops across the two glass panes. First find these and then the temperature drop across the air gap. This problem ignores the increased heat transfer in the air gap due to convection.)

(b) Calculate the rate of heat conduction through a 1.60-cm-thick window of the same area and with the same temperatures. Compare your answer with that for part (a).

Solution:

(a) 83 W

(b) 24 times that of a double pane window.

Exercise:**Problem:**

Many decisions are made on the basis of the payback period: the time it will take through savings to equal the capital cost of an investment. Acceptable payback times depend upon the business or philosophy one has. (For some industries, a payback period is as small as two years.) Suppose you wish to install the extra insulation in [\[link\]](#). If energy cost \$1.00 per million joules and the insulation was \$4.00 per square meter, then calculate the simple payback time. Take the average ΔT for the 120 day heating season to be 15.0°C .

Exercise:**Problem:**

For the human body, what is the rate of heat transfer by conduction through the body's tissue with the following conditions: the tissue thickness is 3.00 cm, the change in temperature is 2.00°C , and the skin area is 1.50 m^2 . How does this compare with the average heat transfer rate to the body resulting from an energy intake of about 2400 kcal per day? (No exercise is included.)

Solution:

20.0 W, 17.2% of 2400 kcal per day

Glossary

R factor

the ratio of thickness to the conductivity of a material

rate of conductive heat transfer

rate of heat transfer from one material to another

thermal conductivity

the property of a material's ability to conduct heat

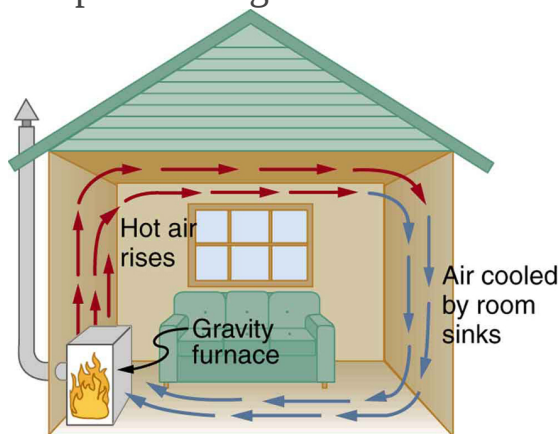
Convection

- Discuss the method of heat transfer by convection.

Convection is driven by large-scale flow of matter. In the case of Earth, the atmospheric circulation is caused by the flow of hot air from the tropics to the poles, and the flow of cold air from the poles toward the tropics. (Note that Earth's rotation causes the observed easterly flow of air in the northern hemisphere). Car engines are kept cool by the flow of water in the cooling system, with the water pump maintaining a flow of cool water to the pistons. The circulatory system is used the body: when the body overheats, the blood vessels in the skin expand (dilate), which increases the blood flow to the skin where it can be cooled by sweating. These vessels become smaller when it is cold outside and larger when it is hot (so more fluid flows, and more energy is transferred).

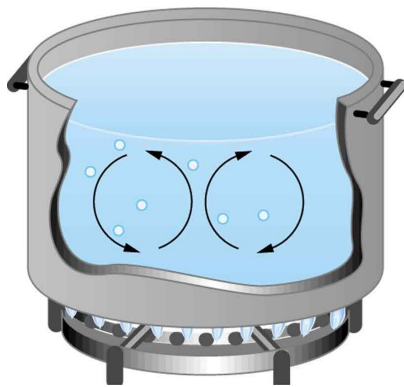
The body also loses a significant fraction of its heat through the breathing process.

While convection is usually more complicated than conduction, we can describe convection and do some straightforward, realistic calculations of its effects. Natural convection is driven by buoyant forces: hot air rises because density decreases as temperature increases. The house in [\[link\]](#) is kept warm in this manner, as is the pot of water on the stove in [\[link\]](#). Ocean currents and large-scale atmospheric circulation transfer energy from one part of the globe to another. Both are examples of natural convection.



Air heated by the so-called

gravity furnace expands and rises, forming a convective loop that transfers energy to other parts of the room. As the air is cooled at the ceiling and outside walls, it contracts, eventually becoming denser than room air and sinking to the floor. A properly designed heating system using natural convection, like this one, can be quite efficient in uniformly heating a home.



Convection plays an important role in heat transfer inside this pot of water. Once conducted to the inside, heat transfer to other parts of the pot is mostly by convection. The hotter water expands, decreases

in density, and rises to transfer heat to other regions of the water, while colder water sinks to the bottom. This process keeps repeating.

Note:

Take-Home Experiment: Convection Rolls in a Heated Pan

Take two small pots of water and use an eye dropper to place a drop of food coloring near the bottom of each. Leave one on a bench top and heat the other over a stovetop. Watch how the color spreads and how long it takes the color to reach the top. Watch how convective loops form.

Example:

Calculating Heat Transfer by Convection: Convection of Air Through the Walls of a House

Most houses are not airtight: air goes in and out around doors and windows, through cracks and crevices, following wiring to switches and outlets, and so on. The air in a typical house is completely replaced in less than an hour. Suppose that a moderately-sized house has inside dimensions $12.0\text{m} \times 18.0\text{m} \times 3.00\text{m}$ high, and that all air is replaced in 30.0 min. Calculate the heat transfer per unit time in watts needed to warm the incoming cold air by 10.0°C , thus replacing the heat transferred by convection alone.

Strategy

Heat is used to raise the temperature of air so that $Q = mc\Delta T$. The rate of heat transfer is then Q/t , where t is the time for air turnover. We are given that ΔT is 10.0°C , but we must still find values for the mass of air and its

specific heat before we can calculate Q . The specific heat of air is a weighted average of the specific heats of nitrogen and oxygen, which gives $c = c_p \cong 1000 \text{ J/kg} \cdot ^\circ \text{C}$ from [\[link\]](#) (note that the specific heat at constant pressure must be used for this process).

Solution

1. Determine the mass of air from its density and the given volume of the house. The density is given from the density ρ and the volume

Equation:

$$m = \rho V = (1.29 \text{ kg/m}^3)(12.0 \text{ m} \times 18.0 \text{ m} \times 3.00 \text{ m}) = 836 \text{ kg}.$$

2. Calculate the heat transferred from the change in air temperature:

$$Q = mc\Delta T \text{ so that}$$

Equation:

$$Q = (836 \text{ kg})(1000 \text{ J/kg} \cdot ^\circ \text{C})(10.0^\circ \text{C}) = 8.36 \times 10^6 \text{ J}.$$

3. Calculate the heat transfer from the heat Q and the turnover time t . Since air is turned over in $t = 0.500 \text{ h} = 1800 \text{ s}$, the heat transferred per unit time is

Equation:

$$\frac{Q}{t} = \frac{8.36 \times 10^6 \text{ J}}{1800 \text{ s}} = 4.64 \text{ kW}.$$

Discussion

This rate of heat transfer is equal to the power consumed by about forty-six 100-W light bulbs. Newly constructed homes are designed for a turnover time of 2 hours or more, rather than 30 minutes for the house of this example. Weather stripping, caulking, and improved window seals are commonly employed. More extreme measures are sometimes taken in very cold (or hot) climates to achieve a tight standard of more than 6 hours for one air turnover. Still longer turnover times are unhealthy, because a minimum amount of fresh air is necessary to supply oxygen for breathing and to dilute household pollutants. The term used for the process by which

outside air leaks into the house from cracks around windows, doors, and the foundation is called “air infiltration.”

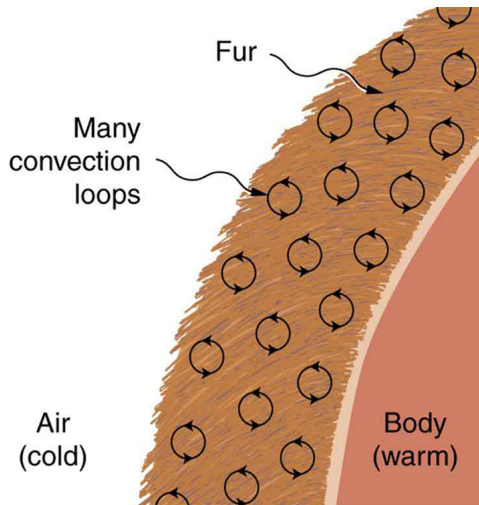
A cold wind is much more chilling than still cold air, because convection combines with conduction in the body to increase the rate at which energy is transferred away from the body. The table below gives approximate wind-chill factors, which are the temperatures of still air that produce the same rate of cooling as air of a given temperature and speed. Wind-chill factors are a dramatic reminder of convection’s ability to transfer heat faster than conduction. For example, a 15.0 m/s wind at 0°C has the chilling equivalent of still air at about -18°C .

Moving air temperature	Wind speed (m/s)				
(°C)	2	5	10	15	20
5	3	−1	−8	−10	−12
2	0	−7	−12	−16	−18
0	−2	−9	−15	−18	−20

Moving air temperature	Wind speed (m/s)				
−5	−7	−15	−22	−26	−29
−10	−12	−21	−29	−34	−36
−20	−23	−34	−44	−50	−52
−40	−44	−59	−73	−82	−84

Wind-Chill Factors

Although air can transfer heat rapidly by convection, it is a poor conductor and thus a good insulator. The amount of available space for airflow determines whether air acts as an insulator or conductor. The space between the inside and outside walls of a house, for example, is about 9 cm (3.5 in)—large enough for convection to work effectively. The addition of wall insulation prevents airflow, so heat loss (or gain) is decreased. Similarly, the gap between the two panes of a double-paned window is about 1 cm, which prevents convection and takes advantage of air's low conductivity to prevent greater loss. Fur, fiber, and fiberglass also take advantage of the low conductivity of air by trapping it in spaces too small to support convection, as shown in the figure. Fur and feathers are lightweight and thus ideal for the protection of animals.



Fur is filled with air, breaking it up into many small pockets. Convection is very slow here, because the loops are so small. The low conductivity of air makes fur a very good lightweight insulator.

Some interesting phenomena happen *when convection is accompanied by a phase change*. It allows us to cool off by sweating, even if the temperature of the surrounding air exceeds body temperature. Heat from the skin is required for sweat to evaporate from the skin, but without air flow, the air becomes saturated and evaporation stops. Air flow caused by convection replaces the saturated air by dry air and evaporation continues.

Example:

Calculate the Flow of Mass during Convection: Sweat-Heat Transfer away from the Body

The average person produces heat at the rate of about 120 W when at rest. At what rate must water evaporate from the body to get rid of all this energy? (This evaporation might occur when a person is sitting in the shade and surrounding temperatures are the same as skin temperature, eliminating heat transfer by other methods.)

Strategy

Energy is needed for a phase change ($Q = mL_v$). Thus, the energy loss per unit time is

Equation:

$$\frac{Q}{t} = \frac{mL_v}{t} = 120 \text{ W} = 120 \text{ J/s}.$$

We divide both sides of the equation by L_v to find that the mass evaporated per unit time is

Equation:

$$\frac{m}{t} = \frac{120 \text{ J/s}}{L_v}.$$

Solution

(1) Insert the value of the latent heat from [\[link\]](#), $L_v = 2430 \text{ kJ/kg} = 2430 \text{ J/g}$. This yields

Equation:

$$\frac{m}{t} = \frac{120 \text{ J/s}}{2430 \text{ J/g}} = 0.0494 \text{ g/s} = 2.96 \text{ g/min}.$$

Discussion

Evaporating about 3 g/min seems reasonable. This would be about 180 g (about 7 oz) per hour. If the air is very dry, the sweat may evaporate without even being noticed. A significant amount of evaporation also takes place in the lungs and breathing passages.

Another important example of the combination of phase change and convection occurs when water evaporates from the oceans. Heat is removed

from the ocean when water evaporates. If the water vapor condenses in liquid droplets as clouds form, heat is released in the atmosphere. Thus, there is an overall transfer of heat from the ocean to the atmosphere. This process is the driving power behind thunderheads, those great cumulus clouds that rise as much as 20.0 km into the stratosphere. Water vapor carried in by convection condenses, releasing tremendous amounts of energy. This energy causes the air to expand and rise, where it is colder. More condensation occurs in these colder regions, which in turn drives the cloud even higher. Such a mechanism is called positive feedback, since the process reinforces and accelerates itself. These systems sometimes produce violent storms, with lightning and hail, and constitute the mechanism driving hurricanes.



Cumulus clouds
are caused by
water vapor that
rises because of
convection. The
rise of clouds is
driven by a
positive
feedback
mechanism.
(credit: Mike
Love)



Convection
accompanied by a
phase change
releases the energy
needed to drive this
thunderhead into the
stratosphere. (credit:
Gerardo García
Moretti)



The phase change that
occurs when this
iceberg melts involves
tremendous heat

transfer. (credit:
Dominic Alves)

The movement of icebergs is another example of convection accompanied by a phase change. Suppose an iceberg drifts from Greenland into warmer Atlantic waters. Heat is removed from the warm ocean water when the ice melts and heat is released to the land mass when the iceberg forms on Greenland.

Exercise:

Check Your Understanding

Problem: Explain why using a fan in the summer feels refreshing!

Solution:

Using a fan increases the flow of air: warm air near your body is replaced by cooler air from elsewhere. Convection increases the rate of heat transfer so that moving air “feels” cooler than still air.

Summary

- Convection is heat transfer by the macroscopic movement of mass. Convection can be natural or forced and generally transfers thermal energy faster than conduction. [\[link\]](#) gives wind-chill factors, indicating that moving air has the same chilling effect of much colder stationary air. *Convection that occurs along with a phase change* can transfer energy from cold regions to warm ones.

Conceptual Questions

Exercise:

Problem:

One way to make a fireplace more energy efficient is to have an external air supply for the combustion of its fuel. Another is to have room air circulate around the outside of the fire box and back into the room. Detail the methods of heat transfer involved in each.

Exercise:**Problem:**

On cold, clear nights horses will sleep under the cover of large trees. How does this help them keep warm?

Problems & Exercises**Exercise:****Problem:**

At what wind speed does -10°C air cause the same chill factor as still air at -29°C ?

Solution:

10 m/s

Exercise:**Problem:**

At what temperature does still air cause the same chill factor as -5°C air moving at 15 m/s?

Exercise:

Problem:

The “steam” above a freshly made cup of instant coffee is really water vapor droplets condensing after evaporating from the hot coffee. What is the final temperature of 250 g of hot coffee initially at 90.0°C if 2.00 g evaporates from it? The coffee is in a Styrofoam cup, so other methods of heat transfer can be neglected.

Solution:

85.7°C

Exercise:**Problem:**

(a) How many kilograms of water must evaporate from a 60.0-kg woman to lower her body temperature by 0.750°C ?

(b) Is this a reasonable amount of water to evaporate in the form of perspiration, assuming the relative humidity of the surrounding air is low?

Exercise:**Problem:**

On a hot dry day, evaporation from a lake has just enough heat transfer to balance the $1.00\text{ kW}/\text{m}^2$ of incoming heat from the Sun. What mass of water evaporates in 1.00 h from each square meter? Explicitly show how you follow the steps in the [Problem-Solving Strategies for the Effects of Heat Transfer](#).

Solution:

1.48 kg

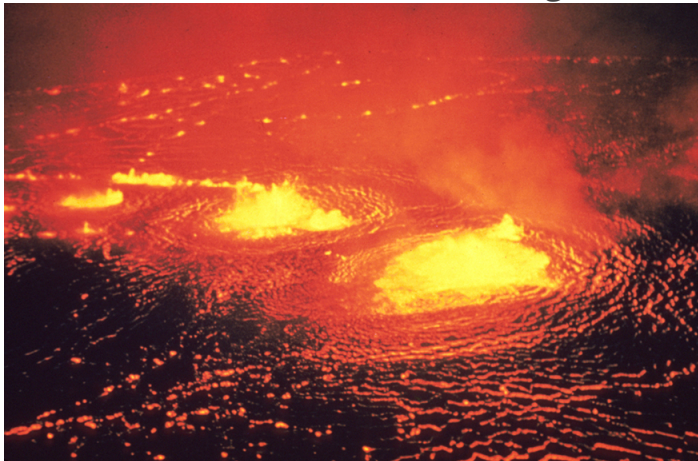
Exercise:

Problem:

One winter day, the climate control system of a large university classroom building malfunctions. As a result, 500 m^3 of excess cold air is brought in each minute. At what rate in kilowatts must heat transfer occur to warm this air by 10.0°C (that is, to bring the air to room temperature)?

Exercise:**Problem:**

The Kilauea volcano in Hawaii is the world's most active, disgorging about $5 \times 10^5 \text{ m}^3$ of 1200°C lava per day. What is the rate of heat transfer out of Earth by convection if this lava has a density of 2700 kg/m^3 and eventually cools to 30°C ? Assume that the specific heat of lava is the same as that of granite.



Lava flow on Kilauea volcano in Hawaii. (credit: J. P. Eaton, U.S. Geological Survey)

Solution:

$$2 \times 10^4 \text{ MW}$$

Exercise:**Problem:**

During heavy exercise, the body pumps 2.00 L of blood per minute to the surface, where it is cooled by 2.00°C . What is the rate of heat transfer from this forced convection alone, assuming blood has the same specific heat as water and its density is 1050 kg/m^3 ?

Exercise:**Problem:**

A person inhales and exhales 2.00 L of 37.0°C air, evaporating $4.00 \times 10^{-2}\text{ g}$ of water from the lungs and breathing passages with each breath.

- (a) How much heat transfer occurs due to evaporation in each breath?
- (b) What is the rate of heat transfer in watts if the person is breathing at a moderate rate of 18.0 breaths per minute?
- (c) If the inhaled air had a temperature of 20.0°C , what is the rate of heat transfer for warming the air?
- (d) Discuss the total rate of heat transfer as it relates to typical metabolic rates. Will this breathing be a major form of heat transfer for this person?

Solution:

- (a) 97.2 J
- (b) 29.2 W
- (c) 9.49 W
- (d) The total rate of heat loss would be $29.2\text{ W} + 9.49\text{ W} = 38.7\text{ W}$. While sleeping, our body consumes 83 W of power, while sitting it

consumes 120 to 210 W. Therefore, the total rate of heat loss from breathing will not be a major form of heat loss for this person.

Exercise:

Problem:

A glass coffee pot has a circular bottom with a 9.00-cm diameter in contact with a heating element that keeps the coffee warm with a continuous heat transfer rate of 50.0 W

(a) What is the temperature of the bottom of the pot, if it is 3.00 mm thick and the inside temperature is 60.0°C?

(b) If the temperature of the coffee remains constant and all of the heat transfer is removed by evaporation, how many grams per minute evaporate? Take the heat of vaporization to be 2340 kJ/kg.

Radiation

- Discuss heat transfer by radiation.
- Explain the power of different materials.

You can feel the heat transfer from a fire and from the Sun. Similarly, you can sometimes tell that the oven is hot without touching its door or looking inside—it may just warm you as you walk by. The space between the Earth and the Sun is largely empty, without any possibility of heat transfer by convection or conduction. In these examples, heat is transferred by radiation. That is, the hot body emits electromagnetic waves that are absorbed by our skin: no medium is required for electromagnetic waves to propagate. Different names are used for electromagnetic waves of different wavelengths: radio waves, microwaves, infrared **radiation**, visible light, ultraviolet radiation, X-rays, and gamma rays.

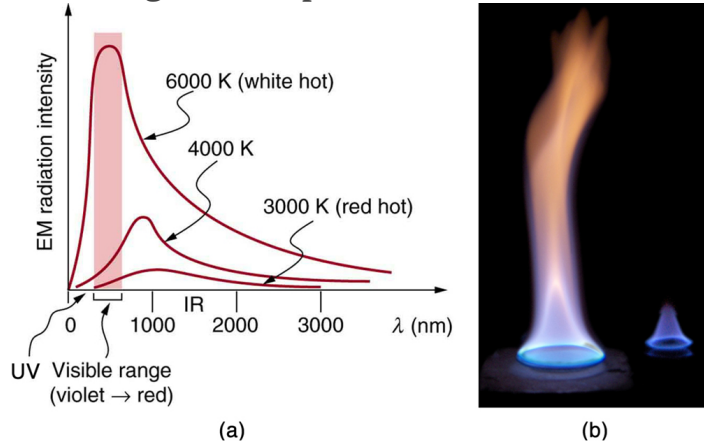


Most of the heat transfer from this fire to the observers is through infrared radiation. The visible light, although dramatic, transfers relatively little thermal energy. Convection transfers energy away from the observers as hot air rises, while conduction is negligibly slow here. Skin is very sensitive to infrared radiation, so that you can sense the presence of a fire

without looking at it directly. (credit:
Daniel X. O'Neil)

The energy of electromagnetic radiation depends on the wavelength (color) and varies over a wide range: a smaller wavelength (or higher frequency) corresponds to a higher energy. Because more heat is radiated at higher temperatures, a temperature change is accompanied by a color change. Take, for example, an electrical element on a stove, which glows from red to orange, while the higher-temperature steel in a blast furnace glows from yellow to white. The radiation you feel is mostly infrared, which corresponds to a lower temperature than that of the electrical element and the steel. The radiated energy depends on its intensity, which is represented in the figure below by the height of the distribution.

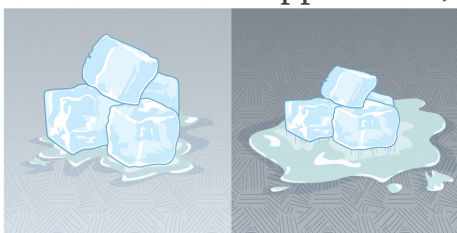
[Electromagnetic Waves](#) explains more about the electromagnetic spectrum and [Introduction to Quantum Physics](#) discusses how the decrease in wavelength corresponds to an increase in energy.



(a) A graph of the spectra of electromagnetic waves emitted from an ideal radiator at three different temperatures. The intensity or rate of radiation emission increases dramatically with temperature, and the spectrum shifts toward the visible and ultraviolet parts of the spectrum. The

shaded portion denotes the visible part of the spectrum. It is apparent that the shift toward the ultraviolet with temperature makes the visible appearance shift from red to white to blue as temperature increases. (b)
Note the variations in color corresponding to variations in flame temperature. (credit: Tuohirulla)

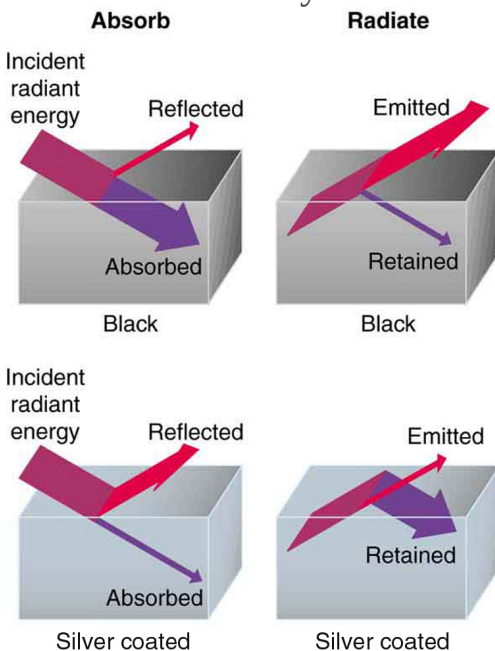
All objects absorb and emit electromagnetic radiation. The rate of heat transfer by radiation is largely determined by the color of the object. Black is the most effective, and white is the least effective. People living in hot climates generally avoid wearing black clothing, for instance (see [\[link\]](#)). Similarly, black asphalt in a parking lot will be hotter than adjacent gray sidewalk on a summer day, because black absorbs better than gray. The reverse is also true—black radiates better than gray. Thus, on a clear summer night, the asphalt will be colder than the gray sidewalk, because black radiates the energy more rapidly than gray. An *ideal radiator* is the same color as an *ideal absorber*, and captures all the radiation that falls on it. In contrast, white is a poor absorber and is also a poor radiator. A white object reflects all radiation, like a mirror. (A perfect, polished white surface is mirror-like in appearance, and a crushed mirror looks white.)



This illustration shows that the darker pavement is hotter than the lighter pavement (much more of the ice on the right has

melted), although both have been in the sunlight for the same time. The thermal conductivities of the pavements are the same.

Gray objects have a uniform ability to absorb all parts of the electromagnetic spectrum. Colored objects behave in similar but more complex ways, which gives them a particular color in the visible range and may make them special in other ranges of the nonvisible spectrum. Take, for example, the strong absorption of infrared radiation by the skin, which allows us to be very sensitive to it.



A black object is a good absorber and a good radiator, while a white (or silver) object is a poor absorber and a poor radiator. It is as if

radiation from the inside is reflected back into the silver object, whereas radiation from the inside of the black object is “absorbed” when it hits the surface and finds itself on the outside and is strongly emitted.

The rate of heat transfer by emitted radiation is determined by the **Stefan-Boltzmann law of radiation**:

Equation:

$$\frac{Q}{t} = \sigma e A T^4,$$

where $\sigma = 5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant, A is the surface area of the object, and T is its absolute temperature in kelvin. The symbol e stands for the **emissivity** of the object, which is a measure of how well it radiates. An ideal jet-black (or black body) radiator has $e = 1$, whereas a perfect reflector has $e = 0$. Real objects fall between these two values. Take, for example, tungsten light bulb filaments which have an e of about 0.5, and carbon black (a material used in printer toner), which has the (greatest known) emissivity of about 0.99.

The radiation rate is directly proportional to the *fourth power* of the absolute temperature—a remarkably strong temperature dependence. Furthermore, the radiated heat is proportional to the surface area of the object. If you knock apart the coals of a fire, there is a noticeable increase in radiation due to an increase in radiating surface area.



A thermograph of part of a building shows temperature variations, indicating where heat transfer to the outside is most severe. Windows are a major region of heat transfer to the outside of homes. (credit: U.S. Army)

Skin is a remarkably good absorber and emitter of infrared radiation, having an emissivity of 0.97 in the infrared spectrum. Thus, we are all nearly (jet) black in the infrared, in spite of the obvious variations in skin color. This high infrared emissivity is why we can so easily feel radiation on our skin. It is also the basis for the use of night scopes used by law enforcement and the military to detect human beings. Even small temperature variations can be detected because of the T^4 dependence. Images, called *thermographs*, can be used medically to detect regions of abnormally high temperature in the body, perhaps indicative of disease. Similar techniques can be used to detect heat leaks in homes [\[link\]](#), optimize performance of blast furnaces, improve comfort levels in work environments, and even remotely map the Earth's temperature profile.

All objects emit and absorb radiation. The *net* rate of heat transfer by radiation (absorption minus emission) is related to both the temperature of the object and the temperature of its surroundings. Assuming that an object

with a temperature T_1 is surrounded by an environment with uniform temperature T_2 , the **net rate of heat transfer by radiation** is

Equation:

$$\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4),$$

where e is the emissivity of the object alone. In other words, it does not matter whether the surroundings are white, gray, or black; the balance of radiation into and out of the object depends on how well it emits and absorbs radiation. When $T_2 > T_1$, the quantity Q_{net}/t is positive; that is, the net heat transfer is from hot to cold.

Note:

Take-Home Experiment: Temperature in the Sun

Place a thermometer out in the sunshine and shield it from direct sunlight using an aluminum foil. What is the reading? Now remove the shield, and note what the thermometer reads. Take a handkerchief soaked in nail polish remover, wrap it around the thermometer and place it in the sunshine. What does the thermometer read?

Example:

Calculate the Net Heat Transfer of a Person: Heat Transfer by Radiation

What is the rate of heat transfer by radiation, with an unclothed person standing in a dark room whose ambient temperature is 22.0°C . The person has a normal skin temperature of 33.0°C and a surface area of 1.50 m^2 . The emissivity of skin is 0.97 in the infrared, where the radiation takes place.

Strategy

We can solve this by using the equation for the rate of radiative heat transfer.

Solution

Insert the temperatures values $T_2 = 295 \text{ K}$ and $T_1 = 306 \text{ K}$, so that

Equation:

$$\frac{Q}{t} = \sigma e A (T_2^4 - T_1^4)$$

Equation:

$$= (5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4)(0.97)(1.50 \text{ m}^2) \left[(295 \text{ K})^4 - (306 \text{ K})^4 \right]$$

Equation:

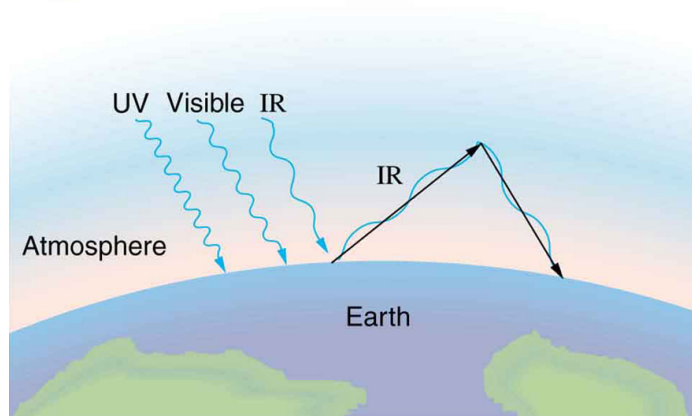
$$= -99 \text{ J/s} = -99 \text{ W}.$$

Discussion

This value is a significant rate of heat transfer to the environment (note the minus sign), considering that a person at rest may produce energy at the rate of 125 W and that conduction and convection will also be transferring energy to the environment. Indeed, we would probably expect this person to feel cold. Clothing significantly reduces heat transfer to the environment by many methods, because clothing slows down both conduction and convection, and has a lower emissivity (especially if it is white) than skin.

The Earth receives almost all its energy from radiation of the Sun and reflects some of it back into outer space. Because the Sun is hotter than the Earth, the net energy flux is from the Sun to the Earth. However, the rate of energy transfer is less than the equation for the radiative heat transfer would predict because the Sun does not fill the sky. The average emissivity (e) of the Earth is about 0.65, but the calculation of this value is complicated by the fact that the highly reflective cloud coverage varies greatly from day to day. There is a negative feedback (one in which a change produces an effect that opposes that change) between clouds and heat transfer; greater temperatures evaporate more water to form more clouds, which reflect more radiation back into space, reducing the temperature. The often mentioned **greenhouse effect** is directly related to the variation of the Earth's emissivity with radiation type (see the figure given below). The greenhouse

effect is a natural phenomenon responsible for providing temperatures suitable for life on Earth. The Earth's relatively constant temperature is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by carbon dioxide (CO_2) and water (H_2O) in the atmosphere and then re-radiated back to the Earth or into outer space. Re-radiation back to the Earth maintains its surface temperature about 40°C higher than it would be if there was no atmosphere, similar to the way glass increases temperatures in a greenhouse.

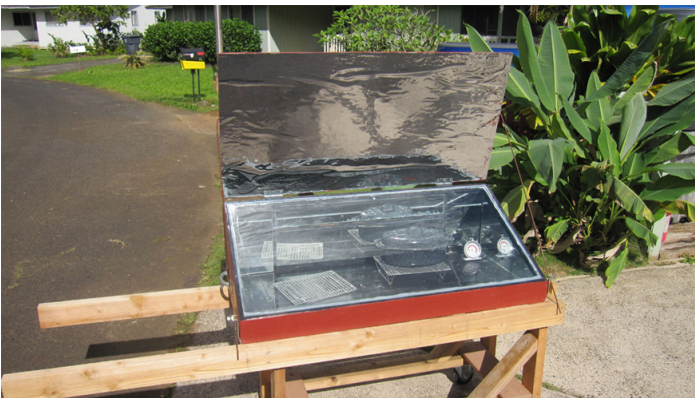


The greenhouse effect is a name given to the trapping of energy in the Earth's atmosphere by a process similar to that used in greenhouses. The atmosphere, like window glass, is transparent to incoming visible radiation and most of the Sun's infrared. These wavelengths are absorbed by the Earth and re-emitted as infrared. Since Earth's temperature is much lower than that of the Sun, the infrared radiated by the Earth has a much longer wavelength. The atmosphere, like glass, traps these longer infrared rays, keeping the Earth warmer than it would otherwise

be. The amount of trapping depends on concentrations of trace gases like carbon dioxide, and a change in the concentration of these gases is believed to affect the Earth's surface temperature.

The greenhouse effect is also central to the discussion of global warming due to emission of carbon dioxide and methane (and other so-called greenhouse gases) into the Earth's atmosphere from industrial production and farming. Changes in global climate could lead to more intense storms, precipitation changes (affecting agriculture), reduction in rain forest biodiversity, and rising sea levels.

Heating and cooling are often significant contributors to energy use in individual homes. Current research efforts into developing environmentally friendly homes quite often focus on reducing conventional heating and cooling through better building materials, strategically positioning windows to optimize radiation gain from the Sun, and opening spaces to allow convection. It is possible to build a zero-energy house that allows for comfortable living in most parts of the United States with hot and humid summers and cold winters.



This simple but effective solar cooker uses the greenhouse effect and reflective material to trap and retain solar energy. Made of inexpensive,

durable materials, it saves money and labor, and is of particular economic value in energy-poor developing countries. (credit: E.B. Kauai)

Conversely, dark space is very cold, about 3K (-454°F), so that the Earth radiates energy into the dark sky. Owing to the fact that clouds have lower emissivity than either oceans or land masses, they reflect some of the radiation back to the surface, greatly reducing heat transfer into dark space, just as they greatly reduce heat transfer into the atmosphere during the day. The rate of heat transfer from soil and grasses can be so rapid that frost may occur on clear summer evenings, even in warm latitudes.

Exercise:

Check Your Understanding

Problem:

What is the change in the rate of the radiated heat by a body at the temperature $T_1 = 20^{\circ}\text{C}$ compared to when the body is at the temperature $T_2 = 40^{\circ}\text{C}$?

Solution:

The radiated heat is proportional to the fourth power of the *absolute temperature*. Because $T_1 = 293\text{ K}$ and $T_2 = 313\text{ K}$, the rate of heat transfer increases by about 30 percent of the original rate.

Note:

Career Connection: Energy Conservation Consultation

The cost of energy is generally believed to remain very high for the foreseeable future. Thus, passive control of heat loss in both commercial and domestic housing will become increasingly important. Energy consultants measure and analyze the flow of energy into and out of houses

and ensure that a healthy exchange of air is maintained inside the house. The job prospects for an energy consultant are strong.

Note:

Problem-Solving Strategies for the Methods of Heat Transfer

1. *Examine the situation to determine what type of heat transfer is involved.*
2. *Identify the type(s) of heat transfer—conduction, convection, or radiation.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is very useful.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. For conduction, equation $\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}$ is appropriate. [\[link\]](#) lists thermal conductivities. For convection, determine the amount of matter moved and use equation $Q = mc\Delta T$, to calculate the heat transfer involved in the temperature change of the fluid. If a phase change accompanies convection, equation $Q = mL_f$ or $Q = mL_v$ is appropriate to find the heat transfer involved in the phase change. [\[link\]](#) lists information relevant to phase change. For radiation, equation $\frac{Q_{\text{net}}}{t} = \sigma eA(T_2^4 - T_1^4)$ gives the net heat transfer rate.
7. *Insert the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.*
8. *Check the answer to see if it is reasonable. Does it make sense?*

Summary

- Radiation is the rate of heat transfer through the emission or absorption of electromagnetic waves.

- The rate of heat transfer depends on the surface area and the fourth power of the absolute temperature:

Equation:

$$\frac{Q}{t} = \sigma e A T^4,$$

where $\sigma = 5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant and e is the emissivity of the body. For a black body, $e = 1$ whereas a shiny white or perfect reflector has $e = 0$, with real objects having values of e between 1 and 0. The net rate of heat transfer by radiation is

Equation:

$$\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4)$$

where T_1 is the temperature of an object surrounded by an environment with uniform temperature T_2 and e is the emissivity of the *object*.

Conceptual Questions

Exercise:

Problem:

When watching a daytime circus in a large, dark-colored tent, you sense significant heat transfer from the tent. Explain why this occurs.

Exercise:

Problem:

Satellites designed to observe the radiation from cold (3 K) dark space have sensors that are shaded from the Sun, Earth, and Moon and that are cooled to very low temperatures. Why must the sensors be at low temperature?

Exercise:

Problem: Why are cloudy nights generally warmer than clear ones?

Exercise:**Problem:**

Why are thermometers that are used in weather stations shielded from the sunshine? What does a thermometer measure if it is shielded from the sunshine and also if it is not?

Exercise:**Problem:**

On average, would Earth be warmer or cooler without the atmosphere? Explain your answer.

Problems & Exercises**Exercise:****Problem:**

At what net rate does heat radiate from a 275-m^2 black roof on a night when the roof's temperature is 30.0°C and the surrounding temperature is 15.0°C ? The emissivity of the roof is 0.900.

Solution:

-21.7 kW

Note that the negative answer implies heat loss to the surroundings.

Exercise:

Problem:

(a) Cherry-red embers in a fireplace are at 850°C and have an exposed area of 0.200 m^2 and an emissivity of 0.980. The surrounding room has a temperature of 18.0°C . If 50% of the radiant energy enters the room, what is the net rate of radiant heat transfer in kilowatts? (b) Does your answer support the contention that most of the heat transfer into a room by a fireplace comes from infrared radiation?

Exercise:**Problem:**

Radiation makes it impossible to stand close to a hot lava flow. Calculate the rate of heat transfer by radiation from 1.00 m^2 of 1200°C fresh lava into 30.0°C surroundings, assuming lava's emissivity is 1.00.

Solution:

−266 kW

Exercise:**Problem:**

(a) Calculate the rate of heat transfer by radiation from a car radiator at 110°C into a 50.0°C environment, if the radiator has an emissivity of 0.750 and a 1.20-m^2 surface area. (b) Is this a significant fraction of the heat transfer by an automobile engine? To answer this, assume a horsepower of 200 hp (1.5 kW) and the efficiency of automobile engines as 25%.

Exercise:

Problem:

Find the net rate of heat transfer by radiation from a skier standing in the shade, given the following. She is completely clothed in white (head to foot, including a ski mask), the clothes have an emissivity of 0.200 and a surface temperature of 10.0°C , the surroundings are at -15.0°C , and her surface area is 1.60 m^2 .

Solution:

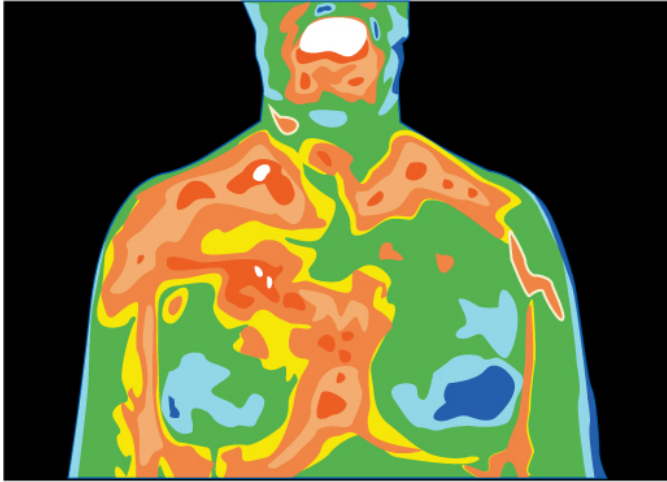
-36.0 W

Exercise:**Problem:**

Suppose you walk into a sauna that has an ambient temperature of 50.0°C . (a) Calculate the rate of heat transfer to you by radiation given your skin temperature is 37.0°C , the emissivity of skin is 0.98, and the surface area of your body is 1.50 m^2 . (b) If all other forms of heat transfer are balanced (the net heat transfer is zero), at what rate will your body temperature increase if your mass is 75.0 kg ?

Exercise:**Problem:**

Thermography is a technique for measuring radiant heat and detecting variations in surface temperatures that may be medically, environmentally, or militarily meaningful. (a) What is the percent increase in the rate of heat transfer by radiation from a given area at a temperature of 34.0°C compared with that at 33.0°C , such as on a person's skin? (b) What is the percent increase in the rate of heat transfer by radiation from a given area at a temperature of 34.0°C compared with that at 20.0°C , such as for warm and cool automobile hoods?



Artist's rendition of a thermograph of a patient's upper body, showing the distribution of heat represented by different colors.

Solution:

(a) 1.31%

(b) 20.5%

Exercise:

Problem:

The Sun radiates like a perfect black body with an emissivity of exactly 1. (a) Calculate the surface temperature of the Sun, given that it is a sphere with a 7.00×10^8 -m radius that radiates 3.80×10^{26} W into 3-K space. (b) How much power does the Sun radiate per square meter of its surface? (c) How much power in watts per square meter is that value at the distance of Earth, 1.50×10^{11} m away? (This number is called the solar constant.)

Exercise:

Problem:

A large body of lava from a volcano has stopped flowing and is slowly cooling. The interior of the lava is at 1200°C , its surface is at 450°C , and the surroundings are at 27.0°C . (a) Calculate the rate at which energy is transferred by radiation from 1.00 m^2 of surface lava into the surroundings, assuming the emissivity is 1.00. (b) Suppose heat conduction to the surface occurs at the same rate. What is the thickness of the lava between the 450°C surface and the 1200°C interior, assuming that the lava's conductivity is the same as that of brick?

Solution:

(a) -15.0 kW

(b) 4.2 cm

Exercise:**Problem:**

Calculate the temperature the entire sky would have to be in order to transfer energy by radiation at 1000 W/m^2 —about the rate at which the Sun radiates when it is directly overhead on a clear day. This value is the effective temperature of the sky, a kind of average that takes account of the fact that the Sun occupies only a small part of the sky but is much hotter than the rest. Assume that the body receiving the energy has a temperature of 27.0°C .

Exercise:

Problem:

(a) A shirtless rider under a circus tent feels the heat radiating from the sunlit portion of the tent. Calculate the temperature of the tent canvas based on the following information: The shirtless rider's skin temperature is 34.0°C and has an emissivity of 0.970. The exposed area of skin is 0.400 m^2 . He receives radiation at the rate of 20.0 W —half what you would calculate if the entire region behind him was hot. The rest of the surroundings are at 34.0°C . (b) Discuss how this situation would change if the sunlit side of the tent was nearly pure white and if the rider was covered by a white tunic.

Solution:

(a) 48.5°C

(b) A pure white object reflects more of the radiant energy that hits it, so a white tent would prevent more of the sunlight from heating up the inside of the tent, and the white tunic would prevent that heat which entered the tent from heating the rider. Therefore, with a white tent, the temperature would be lower than 48.5°C , and the rate of radiant heat transferred to the rider would be less than 20.0 W .

Exercise:**Problem: Integrated Concepts**

One 30.0°C day the relative humidity is 75.0%, and that evening the temperature drops to 20.0°C , well below the dew point. (a) How many grams of water condense from each cubic meter of air? (b) How much heat transfer occurs by this condensation? (c) What temperature increase could this cause in dry air?

Exercise:**Problem: Integrated Concepts**

Large meteors sometimes strike the Earth, converting most of their kinetic energy into thermal energy. (a) What is the kinetic energy of a 10^9 kg meteor moving at 25.0 km/s? (b) If this meteor lands in a deep ocean and 80% of its kinetic energy goes into heating water, how many kilograms of water could it raise by 5.0°C ? (c) Discuss how the energy of the meteor is more likely to be deposited in the ocean and the likely effects of that energy.

Solution:

(a) 3×10^{17} J

(b) 1×10^{13} kg

(c) When a large meteor hits the ocean, it causes great tidal waves, dissipating large amount of its energy in the form of kinetic energy of the water.

Exercise:

Problem: Integrated Concepts

Frozen waste from airplane toilets has sometimes been accidentally ejected at high altitude. Ordinarily it breaks up and disperses over a large area, but sometimes it holds together and strikes the ground. Calculate the mass of 0°C ice that can be melted by the conversion of kinetic and gravitational potential energy when a 20.0 kg piece of frozen waste is released at 12.0 km altitude while moving at 250 m/s and strikes the ground at 100 m/s (since less than 20.0 kg melts, a significant mess results).

Exercise:

Problem: Integrated Concepts

(a) A large electrical power facility produces 1600 MW of “waste heat,” which is dissipated to the environment in cooling towers by warming air flowing through the towers by 5.00°C . What is the

necessary flow rate of air in m^3/s ? (b) Is your result consistent with the large cooling towers used by many large electrical power plants?

Solution:

(a) $3.44 \times 10^5 \text{ m}^3/\text{s}$

(b) This is equivalent to 12 million cubic feet of air per second. That is tremendous. This is too large to be dissipated by heating the air by only 5°C . Many of these cooling towers use the circulation of cooler air over warmer water to increase the rate of evaporation. This would allow much smaller amounts of air necessary to remove such a large amount of heat because evaporation removes larger quantities of heat than was considered in part (a).

Exercise:

Problem: Integrated Concepts

(a) Suppose you start a workout on a Stairmaster, producing power at the same rate as climbing 116 stairs per minute. Assuming your mass is 76.0 kg and your efficiency is 20.0% , how long will it take for your body temperature to rise 1.00°C if all other forms of heat transfer in and out of your body are balanced? (b) Is this consistent with your experience in getting warm while exercising?

Exercise:

Problem: Integrated Concepts

A 76.0-kg person suffering from hypothermia comes indoors and shivers vigorously. How long does it take the heat transfer to increase the person's body temperature by 2.00°C if all other forms of heat transfer are balanced?

Solution:

20.9 min

Exercise:**Problem: Integrated Concepts**

In certain large geographic regions, the underlying rock is hot. Wells can be drilled and water circulated through the rock for heat transfer for the generation of electricity. (a) Calculate the heat transfer that can be extracted by cooling 1.00 km^3 of granite by 100°C . (b) How long will this take if heat is transferred at a rate of 300 MW, assuming no heat transfers back into the 1.00 km of rock by its surroundings?

Exercise:**Problem: Integrated Concepts**

Heat transfers from your lungs and breathing passages by evaporating water. (a) Calculate the maximum number of grams of water that can be evaporated when you inhale 1.50 L of 37°C air with an original relative humidity of 40.0% . (Assume that body temperature is also 37°C .) (b) How many joules of energy are required to evaporate this amount? (c) What is the rate of heat transfer in watts from this method, if you breathe at a normal resting rate of 10.0 breaths per minute?

Solution:

(a) $3.96 \times 10^{-2} \text{ g}$

(b) 96.2 J

(c) 16.0 W

Exercise:**Problem: Integrated Concepts**

(a) What is the temperature increase of water falling 55.0 m over Niagara Falls? (b) What fraction must evaporate to keep the temperature constant?

Exercise:**Problem: Integrated Concepts**

Hot air rises because it has expanded. It then displaces a greater volume of cold air, which increases the buoyant force on it. (a) Calculate the ratio of the buoyant force to the weight of 50.0°C air surrounded by 20.0°C air. (b) What energy is needed to cause 1.00 m^3 of air to go from 20.0°C to 50.0°C ? (c) What gravitational potential energy is gained by this volume of air if it rises 1.00 m ? Will this cause a significant cooling of the air?

Solution:

(a) 1.102

(b) $2.79 \times 10^4\text{ J}$

(c) 12.6 J. This will not cause a significant cooling of the air because it is much less than the energy found in part (b), which is the energy required to warm the air from 20.0°C to 50.0°C .

Exercise:**Problem: Unreasonable Results**

(a) What is the temperature increase of an 80.0 kg person who consumes 2500 kcal of food in one day with 95.0% of the energy transferred as heat to the body? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) 36°C

(b) Any temperature increase greater than about 3°C would be unreasonably large. In this case the final temperature of the person would rise to 73°C (163°F).

(c) The assumption of 95% heat retention is unreasonable.

Exercise:

Problem: Unreasonable Results

A slightly deranged Arctic inventor surrounded by ice thinks it would be much less mechanically complex to cool a car engine by melting ice on it than by having a water-cooled system with a radiator, water pump, antifreeze, and so on. (a) If 80.0% of the energy in 1.00 gal of gasoline is converted into “waste heat” in a car engine, how many kilograms of 0°C ice could it melt? (b) Is this a reasonable amount of ice to carry around to cool the engine for 1.00 gal of gasoline consumption? (c) What premises or assumptions are unreasonable?

Exercise:

Problem: Unreasonable Results

(a) Calculate the rate of heat transfer by conduction through a window with an area of 1.00 m^2 that is 0.750 cm thick, if its inner surface is at 22.0°C and its outer surface is at 35.0°C. (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) 1.46 kW

(b) Very high power loss through a window. An electric heater of this power can keep an entire room warm.

(c) The surface temperatures of the window do not differ by as great an amount as assumed. The inner surface will be warmer, and the outer surface will be cooler.

Exercise:

Problem: Unreasonable Results

A meteorite 1.20 cm in diameter is so hot immediately after penetrating the atmosphere that it radiates 20.0 kW of power. (a) What is its temperature, if the surroundings are at 20.0°C and it has an emissivity of 0.800? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider a new model of commercial airplane having its brakes tested as a part of the initial flight permission procedure. The airplane is brought to takeoff speed and then stopped with the brakes alone. Construct a problem in which you calculate the temperature increase of the brakes during this process. You may assume most of the kinetic energy of the airplane is converted to thermal energy in the brakes and surrounding materials, and that little escapes. Note that the brakes are expected to become so hot in this procedure that they ignite and, in order to pass the test, the airplane must be able to withstand the fire for some time without a general conflagration.

Exercise:

Problem: Construct Your Own Problem

Consider a person outdoors on a cold night. Construct a problem in which you calculate the rate of heat transfer from the person by all three heat transfer methods. Make the initial circumstances such that at rest the person will have a net heat transfer and then decide how much physical activity of a chosen type is necessary to balance the rate of heat transfer. Among the things to consider are the size of the person, type of clothing, initial metabolic rate, sky conditions, amount of water evaporated, and volume of air breathed. Of course, there are many other factors to consider and your instructor may wish to guide you in the assumptions made as well as the detail of analysis and method of presenting your results.

Glossary

emissivity

measure of how well an object radiates

greenhouse effect

warming of the Earth that is due to gases such as carbon dioxide and methane that absorb infrared radiation from the Earth's surface and reradiate it in all directions, thus sending a fraction of it back toward the surface of the Earth

net rate of heat transfer by radiation

is $\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4)$

radiation

energy transferred by electromagnetic waves directly as a result of a temperature difference

Stefan-Boltzmann law of radiation

$\frac{Q}{t} = \sigma e A T^4$ where σ is the Stefan-Boltzmann constant, A is the surface area of the object, T is the absolute temperature, and e is the emissivity

Introduction to Thermodynamics

class="introduction"

A steam
engine
uses heat
transfer
to do
work.

Tourists
regularly
ride this
narrow-
gauge
steam
engine
train near
the San
Juan
Skyway
in
Durango,
Colorado
, part of
the
National
Scenic
Byways
Program.
(credit:
Dennis
Adams)



Heat transfer is energy in transit, and it can be used to do work. It can also be converted to any other form of energy. A car engine, for example, burns fuel for heat transfer into a gas. Work is done by the gas as it exerts a force through a distance, converting its energy into a variety of other forms—into the car's kinetic or gravitational potential energy; into electrical energy to run the spark plugs, radio, and lights; and back into stored energy in the car's battery. But most of the heat transfer produced from burning fuel in the engine does not do work on the gas. Rather, the energy is released into the environment, implying that the engine is quite inefficient.

It is often said that modern gasoline engines cannot be made to be significantly more efficient. We hear the same about heat transfer to electrical energy in large power stations, whether they are coal, oil, natural gas, or nuclear powered. Why is that the case? Is the inefficiency caused by design problems that could be solved with better engineering and superior materials? Is it part of some money-making conspiracy by those who sell energy? Actually, the truth is more interesting, and reveals much about the nature of heat transfer.

Basic physical laws govern how heat transfer for doing work takes place and place insurmountable limits onto its efficiency. This chapter will explore these laws as well as many applications and concepts associated

with them. These topics are part of *thermodynamics*—the study of heat transfer and its relationship to doing work.

The First Law of Thermodynamics

- Define the first law of thermodynamics.
- Describe how conservation of energy relates to the first law of thermodynamics.
- Identify instances of the first law of thermodynamics working in everyday situations, including biological metabolism.
- Calculate changes in the internal energy of a system, after accounting for heat transfer and work done.



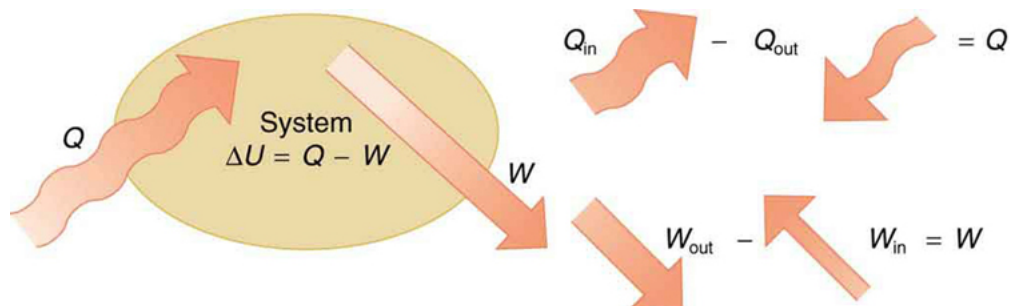
This boiling tea kettle represents energy in motion. The water in the kettle is turning to water vapor because heat is being transferred from the stove to the kettle. As the entire system gets hotter, work is done—from the evaporation of the water to the whistling of the kettle. (credit: Gina Hamilton)

If we are interested in how heat transfer is converted into doing work, then the conservation of energy principle is important. The first law of thermodynamics applies the conservation of energy principle to systems where heat transfer and doing work are the methods of transferring energy into and out of the system. The **first law of thermodynamics** states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system. In equation form, the first law of thermodynamics is

Equation:

$$\Delta U = Q - W.$$

Here ΔU is the *change in internal energy* U of the system. Q is the *net heat transferred into the system*—that is, Q is the sum of all heat transfer into and out of the system. W is the *net work done by the system*—that is, W is the sum of all work done on or by the system. We use the following sign conventions: if Q is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive Q adds energy to the system and positive W takes energy from the system. Thus $\Delta U = Q - W$. Note also that if more heat transfer into the system occurs than work done, the difference is stored as internal energy. Heat engines are a good example of this—heat transfer into them takes place so that they can do work. (See [\[link\]](#).) We will now examine Q , W , and ΔU further.



The first law of thermodynamics is the conservation-of-energy principle stated for a system where heat and work are the methods of transferring energy for a system in

thermal equilibrium. Q represents the net heat transfer—it is the sum of all heat transfers into and out of the system. Q is positive for net heat transfer *into* the system. W is the total work done on and by the system. W is positive when more work is done *by* the system than on it. The change in the internal energy of the system, ΔU , is related to heat and work by the first law of thermodynamics, $\Delta U = Q - W$.

Note:

Making Connections: Law of Thermodynamics and Law of Conservation of Energy

The first law of thermodynamics is actually the law of conservation of energy stated in a form most useful in thermodynamics. The first law gives the relationship between heat transfer, work done, and the change in internal energy of a system.

Heat Q and Work W

Heat transfer (Q) and doing work (W) are the two everyday means of bringing energy into or taking energy out of a system. The processes are quite different. Heat transfer, a less organized process, is driven by temperature differences. Work, a quite organized process, involves a macroscopic force exerted through a distance. Nevertheless, heat and work can produce identical results. For example, both can cause a temperature increase. Heat transfer into a system, such as when the Sun warms the air in a bicycle tire, can increase its temperature, and so can work done on the system, as when the bicyclist pumps air into the tire. Once the temperature increase has occurred, it is impossible to tell whether it was caused by heat transfer or by doing work. This uncertainty is an important point. Heat transfer and work are both energy in transit—neither is stored as such in a

system. However, both can change the internal energy U of a system. Internal energy is a form of energy completely different from either heat or work.

Internal Energy U

We can think about the internal energy of a system in two different but consistent ways. The first is the atomic and molecular view, which examines the system on the atomic and molecular scale. The **internal energy** U of a system is the sum of the kinetic and potential energies of its atoms and molecules. Recall that kinetic plus potential energy is called mechanical energy. Thus internal energy is the sum of atomic and molecular mechanical energy. Because it is impossible to keep track of all individual atoms and molecules, we must deal with averages and distributions. A second way to view the internal energy of a system is in terms of its macroscopic characteristics, which are very similar to atomic and molecular average values.

Macroscopically, we define the change in internal energy ΔU to be that given by the first law of thermodynamics:

Equation:

$$\Delta U = Q - W.$$

Many detailed experiments have verified that $\Delta U = Q - W$, where ΔU is the change in total kinetic and potential energy of all atoms and molecules in a system. It has also been determined experimentally that the internal energy U of a system depends only on the state of the system and *not how it reached that state*. More specifically, U is found to be a function of a few macroscopic quantities (pressure, volume, and temperature, for example), independent of past history such as whether there has been heat transfer or work done. This independence means that if we know the state of a system, we can calculate changes in its internal energy U from a few macroscopic variables.

Note:**Making Connections: Macroscopic and Microscopic**

In thermodynamics, we often use the macroscopic picture when making calculations of how a system behaves, while the atomic and molecular picture gives underlying explanations in terms of averages and distributions. We shall see this again in later sections of this chapter. For example, in the topic of entropy, calculations will be made using the atomic and molecular view.

To get a better idea of how to think about the internal energy of a system, let us examine a system going from State 1 to State 2. The system has internal energy U_1 in State 1, and it has internal energy U_2 in State 2, no matter how it got to either state. So the change in internal energy $\Delta U = U_2 - U_1$ is independent of what caused the change. In other words, ΔU is *independent of path*. By path, we mean the method of getting from the starting point to the ending point. Why is this independence important? Note that $\Delta U = Q - W$. Both Q and W depend on path, but ΔU does not. This path independence means that internal energy U is easier to consider than either heat transfer or work done.

Example:**Calculating Change in Internal Energy: The Same Change in U is Produced by Two Different Processes**

(a) Suppose there is heat transfer of 40.00 J to a system, while the system does 10.00 J of work. Later, there is heat transfer of 25.00 J out of the system while 4.00 J of work is done on the system. What is the net change in internal energy of the system?

(b) What is the change in internal energy of a system when a total of 150.00 J of heat transfer occurs out of (from) the system and 159.00 J of work is done on the system? (See [\[link\]](#)).

Strategy

In part (a), we must first find the net heat transfer and net work done from the given information. Then the first law of thermodynamics

$(\Delta U = Q - W)$ can be used to find the change in internal energy. In part (b), the net heat transfer and work done are given, so the equation can be used directly.

Solution for (a)

The net heat transfer is the heat transfer into the system minus the heat transfer out of the system, or

Equation:

$$Q = 40.00 \text{ J} - 25.00 \text{ J} = 15.00 \text{ J}.$$

Similarly, the total work is the work done by the system minus the work done on the system, or

Equation:

$$W = 10.00 \text{ J} - 4.00 \text{ J} = 6.00 \text{ J}.$$

Thus the change in internal energy is given by the first law of thermodynamics:

Equation:

$$\Delta U = Q - W = 15.00 \text{ J} - 6.00 \text{ J} = 9.00 \text{ J}.$$

We can also find the change in internal energy for each of the two steps. First, consider 40.00 J of heat transfer in and 10.00 J of work out, or

Equation:

$$\Delta U_1 = Q_1 - W_1 = 40.00 \text{ J} - 10.00 \text{ J} = 30.00 \text{ J}.$$

Now consider 25.00 J of heat transfer out and 4.00 J of work in, or

Equation:

$$\Delta U_2 = Q_2 - W_2 = -25.00 \text{ J} - (-4.00 \text{ J}) = -21.00 \text{ J}.$$

The total change is the sum of these two steps, or

Equation:

$$\Delta U = \Delta U_1 + \Delta U_2 = 30.00 \text{ J} + (-21.00 \text{ J}) = 9.00 \text{ J}.$$

Discussion on (a)

No matter whether you look at the overall process or break it into steps, the change in internal energy is the same.

Solution for (b)

Here the net heat transfer and total work are given directly to be

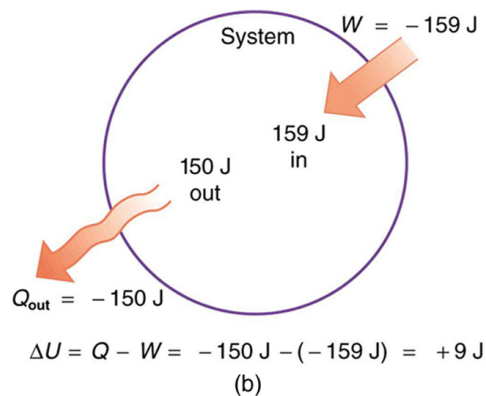
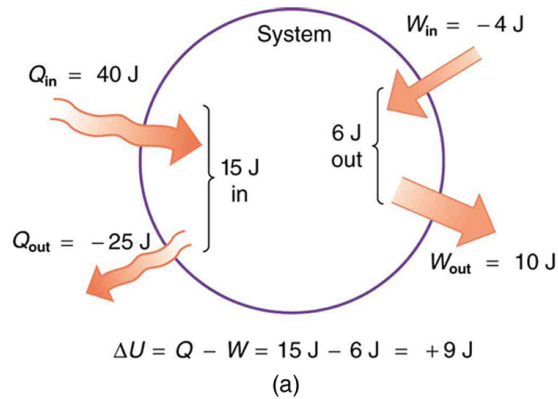
$Q = -150.00 \text{ J}$ and $W = -159.00 \text{ J}$, so that

Equation:

$$\Delta U = Q - W = -150.00 \text{ J} - (-159.00 \text{ J}) = 9.00 \text{ J}.$$

Discussion on (b)

A very different process in part (b) produces the same 9.00-J change in internal energy as in part (a). Note that the change in the system in both parts is related to ΔU and not to the individual Q s or W s involved. The system ends up in the *same* state in both (a) and (b). Parts (a) and (b) present two different paths for the system to follow between the same starting and ending points, and the change in internal energy for each is the same—it is independent of path.



Two different processes produce the same change in a system. (a) A total of 15.00 J of heat transfer occurs into the system, while work takes out a total of 6.00 J. The change in internal energy is $\Delta U = Q - W = 9.00 \text{ J}$. (b) Heat transfer removes 150.00 J from the system while work puts 159.00 J into it, producing an increase of 9.00 J in internal energy. If the system starts out in the same state in (a) and (b), it will end up in the same final state in either case—its final state is related to internal

energy, not how that energy
was acquired.

Human Metabolism and the First Law of Thermodynamics

Human metabolism is the conversion of food into heat transfer, work, and stored fat. Metabolism is an interesting example of the first law of thermodynamics in action. We now take another look at these topics via the first law of thermodynamics. Considering the body as the system of interest, we can use the first law to examine heat transfer, doing work, and internal energy in activities ranging from sleep to heavy exercise. What are some of the major characteristics of heat transfer, doing work, and energy in the body? For one, body temperature is normally kept constant by heat transfer to the surroundings. This means Q is negative. Another fact is that the body usually does work on the outside world. This means W is positive. In such situations, then, the body loses internal energy, since $\Delta U = Q - W$ is negative.

Now consider the effects of eating. Eating increases the internal energy of the body by adding chemical potential energy (this is an unromantic view of a good steak). The body *metabolizes* all the food we consume. Basically, metabolism is an oxidation process in which the chemical potential energy of food is released. This implies that food input is in the form of work. Food energy is reported in a special unit, known as the Calorie. This energy is measured by burning food in a calorimeter, which is how the units are determined.

In chemistry and biochemistry, one calorie (spelled with a *lowercase c*) is defined as the energy (or heat transfer) required to raise the temperature of one gram of pure water by one degree Celsius. Nutritionists and weight-watchers tend to use the *dietary* calorie, which is frequently called a Calorie (spelled with a *capital C*). One food Calorie is the energy needed to raise the temperature of one *kilogram* of water by one degree Celsius. This

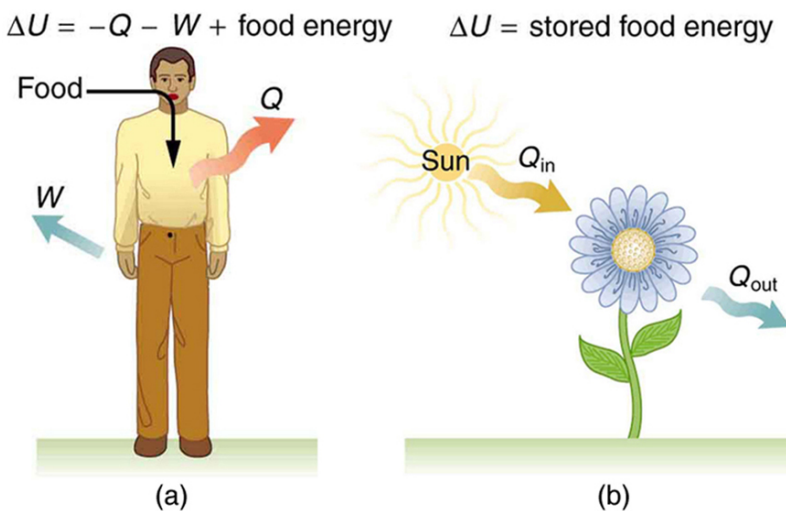
means that one dietary Calorie is equal to one kilocalorie for the chemist, and one must be careful to avoid confusion between the two.

Again, consider the internal energy the body has lost. There are three places this internal energy can go—to heat transfer, to doing work, and to stored fat (a tiny fraction also goes to cell repair and growth). Heat transfer and doing work take internal energy out of the body, and food puts it back. If you eat just the right amount of food, then your average internal energy remains constant. Whatever you lose to heat transfer and doing work is replaced by food, so that, in the long run, $\Delta U = 0$. If you overeat repeatedly, then ΔU is always positive, and your body stores this extra internal energy as fat. The reverse is true if you eat too little. If ΔU is negative for a few days, then the body metabolizes its own fat to maintain body temperature and do work that takes energy from the body. This process is how dieting produces weight loss.

Life is not always this simple, as any dieter knows. The body stores fat or metabolizes it only if energy intake changes for a period of several days. Once you have been on a major diet, the next one is less successful because your body alters the way it responds to low energy intake. Your basal metabolic rate (BMR) is the rate at which food is converted into heat transfer and work done while the body is at complete rest. The body adjusts its basal metabolic rate to partially compensate for over-eating or under-eating. The body will decrease the metabolic rate rather than eliminate its own fat to replace lost food intake. You will chill more easily and feel less energetic as a result of the lower metabolic rate, and you will not lose weight as fast as before. Exercise helps to lose weight, because it produces both heat transfer from your body and work, and raises your metabolic rate even when you are at rest. Weight loss is also aided by the quite low efficiency of the body in converting internal energy to work, so that the loss of internal energy resulting from doing work is much greater than the work done. It should be noted, however, that living systems are not in thermalequilibrium.

The body provides us with an excellent indication that many thermodynamic processes are *irreversible*. An irreversible process can go in one direction but not the reverse, under a given set of conditions. For

example, although body fat can be converted to do work and produce heat transfer, work done on the body and heat transfer into it cannot be converted to body fat. Otherwise, we could skip lunch by sunning ourselves or by walking down stairs. Another example of an irreversible thermodynamic process is photosynthesis. This process is the intake of one form of energy—light—by plants and its conversion to chemical potential energy. Both applications of the first law of thermodynamics are illustrated in [\[link\]](#). One great advantage of conservation laws such as the first law of thermodynamics is that they accurately describe the beginning and ending points of complex processes, such as metabolism and photosynthesis, without regard to the complications in between. [\[link\]](#) presents a summary of terms relevant to the first law of thermodynamics.



(a) The first law of thermodynamics applied to metabolism. Heat transferred out of the body (Q) and work done by the body (W) remove internal energy, while food intake replaces it. (Food intake may be considered as work done on the body.) (b) Plants convert part of the radiant heat transfer in sunlight to stored chemical energy, a process called photosynthesis.

Term	Definition
U	Internal energy—the sum of the kinetic and potential energies of a system’s atoms and molecules. Can be divided into many subcategories, such as thermal and chemical energy. Depends only on the state of a system (such as its P , V , and T), not on how the energy entered the system. Change in internal energy is path independent.
Q	Heat—energy transferred because of a temperature difference. Characterized by random molecular motion. Highly dependent on path. Q entering a system is positive.
W	Work—energy transferred by a force moving through a distance. An organized, orderly process. Path dependent. W done by a system (either against an external force or to increase the volume of the system) is positive.

Summary of Terms for the First Law of Thermodynamics, $\Delta U = Q - W$

Section Summary

- The first law of thermodynamics is given as $\Delta U = Q - W$, where ΔU is the change in internal energy of a system, Q is the net heat transfer (the sum of all heat transfer into and out of the system), and W is the net work done (the sum of all work done on or by the system).
- Both Q and W are energy in transit; only ΔU represents an independent quantity capable of being stored.

- The internal energy U of a system depends only on the state of the system and not how it reached that state.
- Metabolism of living organisms, and photosynthesis of plants, are specialized types of heat transfer, doing work, and internal energy of systems.

Conceptual Questions

Exercise:

Problem:

Describe the photo of the tea kettle at the beginning of this section in terms of heat transfer, work done, and internal energy. How is heat being transferred? What is the work done and what is doing it? How does the kettle maintain its internal energy?

Exercise:

Problem:

The first law of thermodynamics and the conservation of energy, as discussed in [Conservation of Energy](#), are clearly related. How do they differ in the types of energy considered?

Exercise:

Problem:

Heat transfer Q and work done W are always energy in transit, whereas internal energy U is energy stored in a system. Give an example of each type of energy, and state specifically how it is either in transit or resides in a system.

Exercise:

Problem:

How do heat transfer and internal energy differ? In particular, which can be stored as such in a system and which cannot?

Exercise:

Problem:

If you run down some stairs and stop, what happens to your kinetic energy and your initial gravitational potential energy?

Exercise:**Problem:**

Give an explanation of how food energy (calories) can be viewed as molecular potential energy (consistent with the atomic and molecular definition of internal energy).

Exercise:**Problem:**

Identify the type of energy transferred to your body in each of the following as either internal energy, heat transfer, or doing work: (a) basking in sunlight; (b) eating food; (c) riding an elevator to a higher floor.

Problems & Exercises**Exercise:****Problem:**

What is the change in internal energy of a car if you put 12.0 gal of gasoline into its tank? The energy content of gasoline is $1.3 \times 10^8 \text{ J/gal}$. All other factors, such as the car's temperature, are constant.

Solution:

$$1.6 \times 10^9 \text{ J}$$

Exercise:

Problem:

How much heat transfer occurs from a system, if its internal energy decreased by 150 J while it was doing 30.0 J of work?

Exercise:**Problem:**

A system does 1.80×10^8 J of work while 7.50×10^8 J of heat transfer occurs to the environment. What is the change in internal energy of the system assuming no other changes (such as in temperature or by the addition of fuel)?

Solution:

$$-9.30 \times 10^8 \text{ J}$$

Exercise:**Problem:**

What is the change in internal energy of a system which does 4.50×10^5 J of work while 3.00×10^6 J of heat transfer occurs into the system, and 8.00×10^6 J of heat transfer occurs to the environment?

Exercise:**Problem:**

Suppose a woman does 500 J of work and 9500 J of heat transfer occurs into the environment in the process. (a) What is the decrease in her internal energy, assuming no change in temperature or consumption of food? (That is, there is no other energy transfer.) (b) What is her efficiency?

Solution:

(a) -1.0×10^4 J , or -2.39 kcal

(b) 5.00%

Exercise:**Problem:**

(a) How much food energy will a man metabolize in the process of doing 35.0 kJ of work with an efficiency of 5.00%? (b) How much heat transfer occurs to the environment to keep his temperature constant? Explicitly show how you follow the steps in the Problem-Solving Strategy for thermodynamics found in [Problem-Solving Strategies for Thermodynamics](#).

Exercise:**Problem:**

(a) What is the average metabolic rate in watts of a man who metabolizes 10,500 kJ of food energy in one day? (b) What is the maximum amount of work in joules he can do without breaking down fat, assuming a maximum efficiency of 20.0%? (c) Compare his work output with the daily output of a 187-W (0.250-horsepower) motor.

Solution:

(a) 122 W

(b) 2.10×10^6 J

(c) Work done by the motor is 1.61×10^7 J ;thus the motor produces 7.67 times the work done by the man

Exercise:**Problem:**

(a) How long will the energy in a 1470-kJ (350-kcal) cup of yogurt last in a woman doing work at the rate of 150 W with an efficiency of 20.0% (such as in leisurely climbing stairs)? (b) Does the time found in part (a) imply that it is easy to consume more food energy than you can reasonably expect to work off with exercise?

Exercise:

Problem:

(a) A woman climbing the Washington Monument metabolizes 6.00×10^2 kJ of food energy. If her efficiency is 18.0%, how much heat transfer occurs to the environment to keep her temperature constant? (b) Discuss the amount of heat transfer found in (a). Is it consistent with the fact that you quickly warm up when exercising?

Solution:

(a) 492 kJ

(b) This amount of heat is consistent with the fact that you warm quickly when exercising. Since the body is inefficient, the excess heat produced must be dissipated through sweating, breathing, etc.

Glossary

first law of thermodynamics

states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system

internal energy

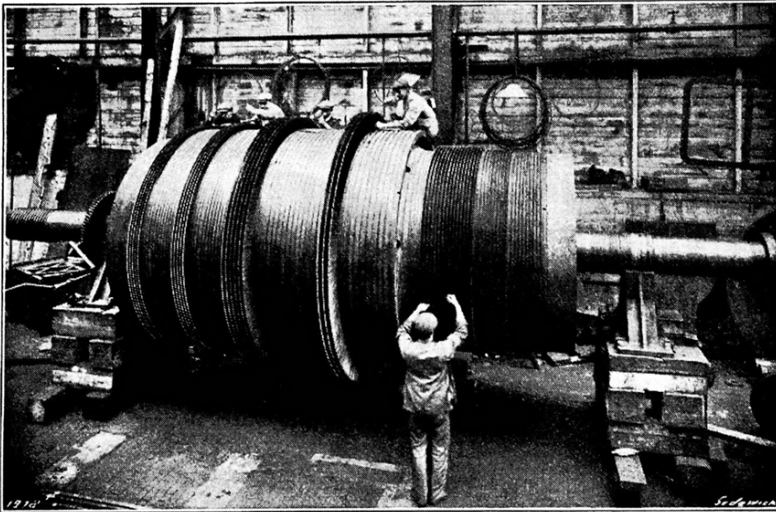
the sum of the kinetic and potential energies of a system's atoms and molecules

human metabolism

conversion of food into heat transfer, work, and stored fat

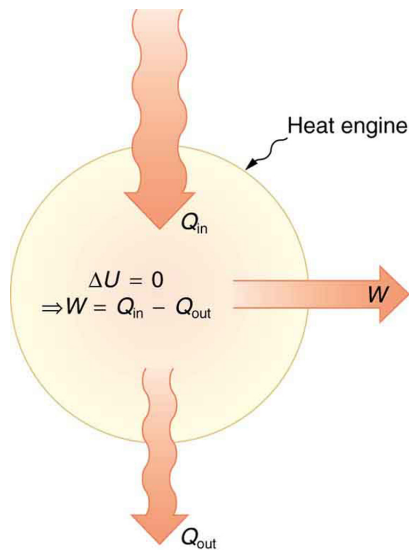
The First Law of Thermodynamics and Some Simple Processes

- Describe the processes of a simple heat engine.
- Explain the differences among the simple thermodynamic processes— isobaric, isochoric, isothermal, and adiabatic.
- Calculate total work done in a cyclical thermodynamic process.

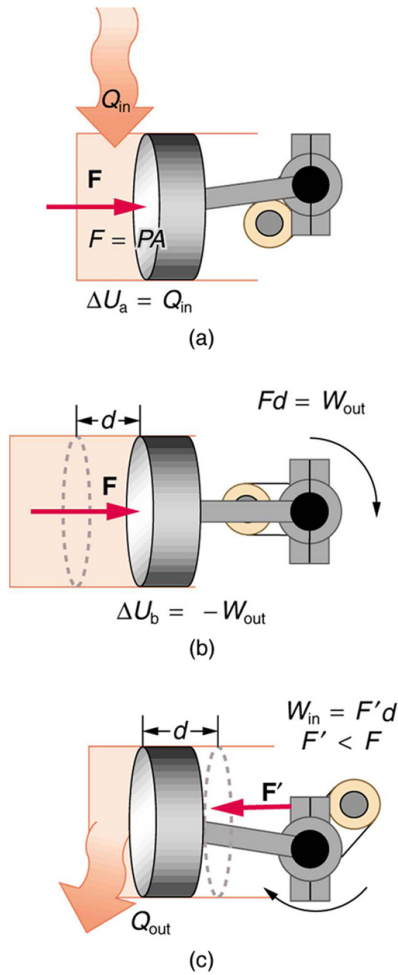


Beginning with the Industrial Revolution, humans have harnessed power through the use of the first law of thermodynamics, before we even understood it completely. This photo, of a steam engine at the Turbinia Works, dates from 1911, a mere 61 years after the first explicit statement of the first law of thermodynamics by Rudolph Clausius. (credit: public domain; author unknown)

One of the most important things we can do with heat transfer is to use it to do work for us. Such a device is called a **heat engine**. Car engines and steam turbines that generate electricity are examples of heat engines. [\[link\]](#) shows schematically how the first law of thermodynamics applies to the typical heat engine.



Schematic representation of a heat engine, governed, of course, by the first law of thermodynamics. It is impossible to devise a system where $Q_{out} = 0$, that is, in which no heat transfer occurs to the environment.



(a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease when it expands, indicating

that the gas's
internal energy has
been decreased by
doing work. (c)
Heat transfer to the
environment
further reduces
pressure in the gas
so that the piston
can be more easily
returned to its
starting position.

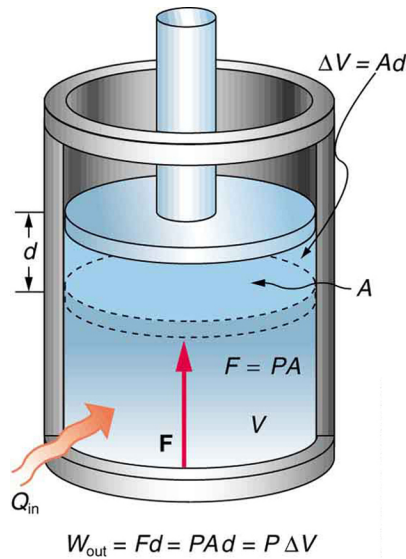
The illustrations above show one of the ways in which heat transfer does work. Fuel combustion produces heat transfer to a gas in a cylinder, increasing the pressure of the gas and thereby the force it exerts on a movable piston. The gas does work on the outside world, as this force moves the piston through some distance. Heat transfer to the gas cylinder results in work being done. To repeat this process, the piston needs to be returned to its starting point. Heat transfer now occurs from the gas to the surroundings so that its pressure decreases, and a force is exerted by the surroundings to push the piston back through some distance. Variations of this process are employed daily in hundreds of millions of heat engines. We will examine heat engines in detail in the next section. In this section, we consider some of the simpler underlying processes on which heat engines are based.

PV Diagrams and their Relationship to Work Done on or by a Gas

A process by which a gas does work on a piston at constant pressure is called an **isobaric process**. Since the pressure is constant, the force exerted is constant and the work done is given as

Equation:

$$P\Delta V.$$



An isobaric expansion of a gas requires heat transfer to keep the pressure constant. Since pressure is constant, the work done is $P\Delta V$.

Equation:

$$W = Fd$$

See the symbols as shown in [\[link\]](#). Now $F = PA$, and so

Equation:

$$W = PAd.$$

Because the volume of a cylinder is its cross-sectional area A times its length d , we see that $Ad = \Delta V$, the change in volume; thus,

Equation:

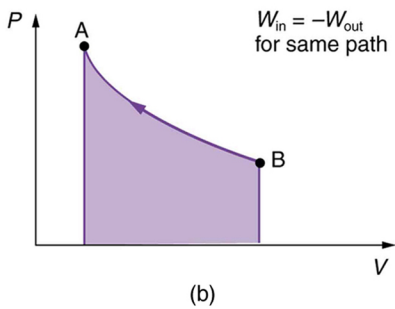
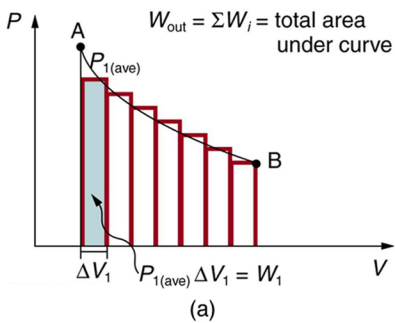
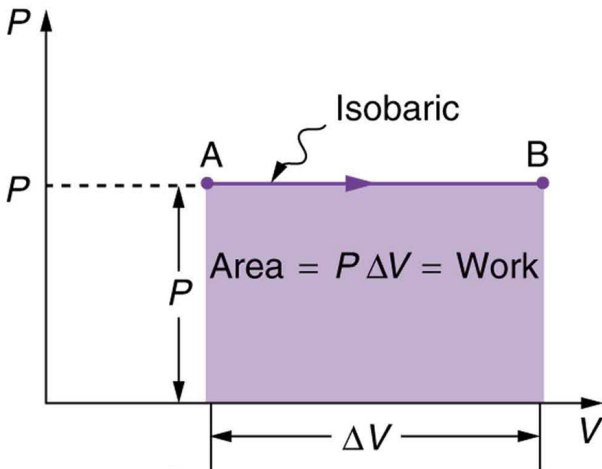
$$W = P\Delta V \text{ (isobaric process).}$$

Note that if ΔV is positive, then W is positive, meaning that work is done by the gas on the outside world.

(Note that the pressure involved in this work that we've called P is the pressure of the gas *inside* the tank. If we call the pressure outside the tank P_{ext} , an expanding gas would be working *against* the external pressure; the work done would therefore be $W = -P_{\text{ext}}\Delta V$ (isobaric process). Many texts use this definition of work, and not the definition based on internal pressure, as the basis of the First Law of Thermodynamics. This definition reverses the sign conventions for work, and results in a statement of the first law that becomes $\Delta U = Q + W$.)

It is not surprising that $W = P\Delta V$, since we have already noted in our treatment of fluids that pressure is a type of potential energy per unit volume and that pressure in fact has units of energy divided by volume. We also noted in our discussion of the ideal gas law that PV has units of energy. In this case, some of the energy associated with pressure becomes work.

[\[link\]](#) shows a graph of pressure versus volume (that is, a PV diagram for an isobaric process. You can see in the figure that the work done is the area under the graph. This property of PV diagrams is very useful and broadly applicable: *the work done on or by a system in going from one state to another equals the area under the curve on a PV diagram.*

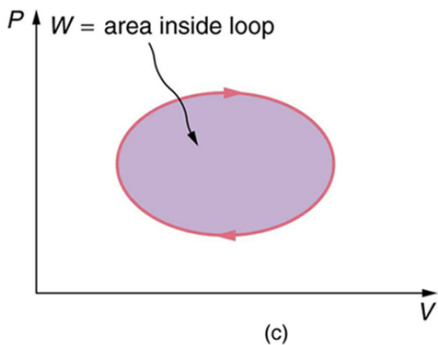
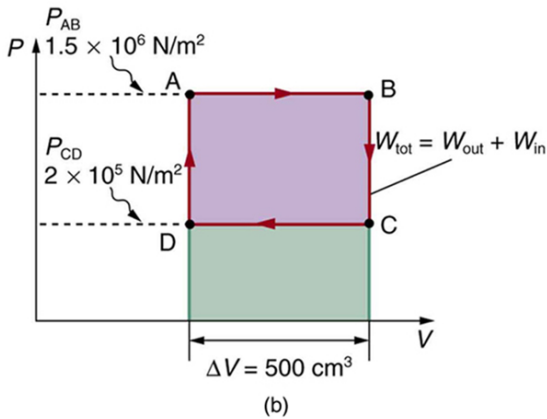
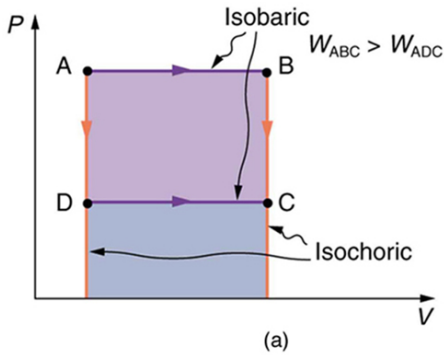


(a) A PV diagram in which pressure varies as well as volume. The work done for each interval is its average pressure times the change in volume, or the area under the curve over that interval. Thus the total area under the curve equals the total work done. (b) Work must be done on the system to follow the reverse path. This is interpreted as a negative area under the curve.

We can see where this leads by considering [\[link\]](#)(a), which shows a more general process in which both pressure and volume change. The area under the curve is closely approximated by dividing it into strips, each having an average constant pressure $P_{i(\text{ave})}$. The work done is $W_i = P_{i(\text{ave})}\Delta V_i$ for each strip, and the total work done is the sum of the W_i . Thus the total work done is the total area under the curve. If the path is reversed, as in [\[link\]](#)(b), then work is done on the system. The area under the curve in that case is negative, because ΔV is negative.

PV diagrams clearly illustrate that *the work done depends on the path taken and not just the endpoints*. This path dependence is seen in [\[link\]](#)(a), where

more work is done in going from A to C by the path via point B than by the path via point D. The vertical paths, where volume is constant, are called **isochoric** processes. Since volume is constant, $\Delta V = 0$, and no work is done in an isochoric process. Now, if the system follows the cyclical path ABCDA, as in [\[link\]](#)(b), then the total work done is the area inside the loop. The negative area below path CD subtracts, leaving only the area inside the rectangle. In fact, the work done in any cyclical process (one that returns to its starting point) is the area inside the loop it forms on a PV diagram, as [\[link\]](#)(c) illustrates for a general cyclical process. Note that the loop must be traversed in the clockwise direction for work to be positive—that is, for there to be a net work output.



(a) The work done in going from A to C depends on path. The work is greater for the path ABC than for the path ADC, because the former is at higher pressure. In both cases, the work done is the area under the path. This area is greater for path ABC. (b) The total work done in the cyclical process

ABCD is the area inside the loop, since the negative area below CD subtracts out, leaving just the area inside the rectangle. (The values given for the pressures and the change in volume are intended for use in the example below.) (c) The area inside any closed loop is the work done in the cyclical process. If the loop is traversed in a clockwise direction, W is positive—it is work done on the outside environment. If the loop is traveled in a counter-clockwise direction, W is negative—it is work that is done to the system.

Example:**Total Work Done in a Cyclical Process Equals the Area Inside the Closed Loop on a PV Diagram**

Calculate the total work done in the cyclical process ABCDA shown in [\[link\]](#)(b) by the following two methods to verify that work equals the area inside the closed loop on the PV diagram. (Take the data in the figure to be precise to three significant figures.) (a) Calculate the work done along each segment of the path and add these values to get the total work. (b) Calculate the area inside the rectangle ABCDA.

Strategy

To find the work along any path on a PV diagram, you use the fact that work is pressure times change in volume, or $W = P\Delta V$. So in part (a),

this value is calculated for each leg of the path around the closed loop.

Solution for (a)

The work along path AB is

Equation:

$$\begin{aligned}W_{AB} &= P_{AB}\Delta V_{AB} \\&= (1.50\times 10^6 \text{ N/m}^2)(5.00\times 10^{-4} \text{ m}^3) = 750 \text{ J.}\end{aligned}$$

Since the path BC is isochoric, $\Delta V_{BC} = 0$, and so $W_{BC} = 0$. The work along path CD is negative, since ΔV_{CD} is negative (the volume decreases). The work is

Equation:

$$\begin{aligned}W_{CD} &= P_{CD}\Delta V_{CD} \\&= (2.00\times 10^5 \text{ N/m}^2)(-5.00\times 10^{-4} \text{ m}^3) = -100 \text{ J.}\end{aligned}$$

Again, since the path DA is isochoric, $\Delta V_{DA} = 0$, and so $W_{DA} = 0$. Now the total work is

Equation:

$$\begin{aligned}W &= W_{AB} + W_{BC} + W_{CD} + W_{DA} \\&= 750 \text{ J} + 0 + (-100\text{J}) + 0 = 650 \text{ J.}\end{aligned}$$

Solution for (b)

The area inside the rectangle is its height times its width, or

Equation:

$$\begin{aligned}\text{area} &= (P_{AB} - P_{CD})\Delta V \\&= \left[(1.50\times 10^6 \text{ N/m}^2) - (2.00\times 10^5 \text{ N/m}^2) \right] (5.00\times 10^{-4} \text{ m}^3) \\&= 650 \text{ J.}\end{aligned}$$

Thus,

Equation:

$$\text{area} = 650 \text{ J} = W.$$

Discussion

The result, as anticipated, is that the area inside the closed loop equals the work done. The area is often easier to calculate than is the work done along each path. It is also convenient to visualize the area inside different curves on PV diagrams in order to see which processes might produce the most work. Recall that work can be done to the system, or by the system, depending on the sign of W . A positive W is work that is done by the system on the outside environment; a negative W represents work done by the environment on the system.

[\[link\]](#)(a) shows two other important processes on a PV diagram. For comparison, both are shown starting from the same point A. The upper curve ending at point B is an **isothermal** process—that is, one in which temperature is kept constant. If the gas behaves like an ideal gas, as is often the case, and if no phase change occurs, then $PV = nRT$. Since T is constant, PV is a constant for an isothermal process. We ordinarily expect the temperature of a gas to decrease as it expands, and so we correctly suspect that heat transfer must occur from the surroundings to the gas to keep the temperature constant during an isothermal expansion. To show this more rigorously for the special case of a monatomic ideal gas, we note that the average kinetic energy of an atom in such a gas is given by

Equation:

$$\frac{1}{2}mv^2 = \frac{3}{2}kT.$$

The kinetic energy of the atoms in a monatomic ideal gas is its only form of internal energy, and so its total internal energy U is

Equation:

$$U = N\frac{1}{2}mv^2 = \frac{3}{2}NkT, \text{ (monatomic ideal gas),}$$

where N is the number of atoms in the gas. This relationship means that the internal energy of an ideal monatomic gas is constant during an isothermal process—that is, $\Delta U = 0$. If the internal energy does not change, then the net heat transfer into the gas must equal the net work done by the gas. That is, because $\Delta U = Q - W = 0$ here, $Q = W$. We must have just enough heat transfer to replace the work done. An isothermal

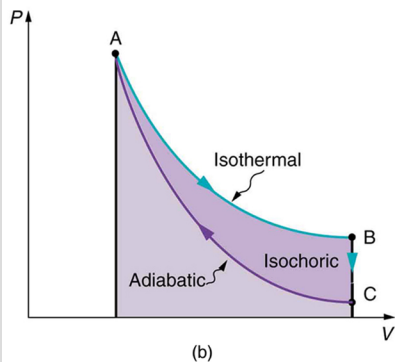
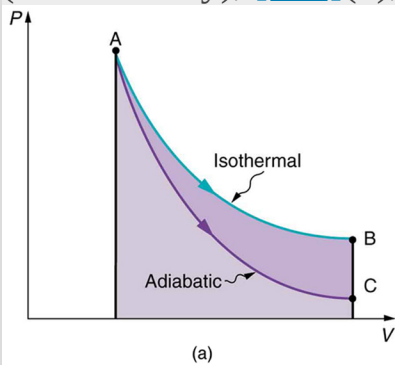
process is inherently slow, because heat transfer occurs continuously to keep the gas temperature constant at all times and must be allowed to spread through the gas so that there are no hot or cold regions.

Also shown in [\[link\]](#)(a) is a curve AC for an **adiabatic** process, defined to be one in which there is no heat transfer—that is, $Q = 0$. Processes that are nearly adiabatic can be achieved either by using very effective insulation or by performing the process so fast that there is little time for heat transfer. Temperature must decrease during an adiabatic expansion process, since work is done at the expense of internal energy:

Equation:

$$U = \frac{3}{2}NkT.$$

(You might have noted that a gas released into atmospheric pressure from a pressurized cylinder is substantially colder than the gas in the cylinder.) In fact, because $Q = 0$, $\Delta U = -W$ for an adiabatic process. Lower temperature results in lower pressure along the way, so that curve AC is lower than curve AB, and less work is done. If the path ABCA could be followed by cooling the gas from B to C at constant volume (isochorically), [\[link\]](#)(b), there would be a net work output.



(a) The upper curve is an isothermal process ($\Delta T = 0$), whereas the lower curve is an adiabatic process ($Q = 0$). Both start from the same point A, but the isothermal process does more work than the adiabatic because heat transfer into the gas takes place to keep its temperature constant. This keeps the pressure higher all along the isothermal path than along the adiabatic path, producing more work. The adiabatic path thus ends up with a lower pressure and temperature at point C, even though the final volume is the same as for the isothermal process.

(b) The cycle ABCA produces a net work output.

Reversible Processes

Both isothermal and adiabatic processes such as shown in [\[link\]](#) are reversible in principle. A **reversible process** is one in which both the system and its environment can return to exactly the states they were in by following the reverse path. The reverse isothermal and adiabatic paths are BA and CA, respectively. Real macroscopic processes are never exactly reversible. In the previous examples, our system is a gas (like that in [\[link\]](#)), and its environment is the piston, cylinder, and the rest of the universe. If there are any energy-dissipating mechanisms, such as friction or turbulence, then heat transfer to the environment occurs for either direction of the piston. So, for example, if the path BA is followed and there is friction, then the gas will be returned to its original state but the environment will not—it will have been heated in both directions. Reversibility requires the direction of heat transfer to reverse for the reverse path. Since dissipative mechanisms cannot be completely eliminated, real processes cannot be reversible.

There must be reasons that real macroscopic processes cannot be reversible. We can imagine them going in reverse. For example, heat transfer occurs spontaneously from hot to cold and never spontaneously the reverse. Yet it would not violate the first law of thermodynamics for this to happen. In fact, all spontaneous processes, such as bubbles bursting, never go in reverse. There is a second thermodynamic law that forbids them from going in reverse. When we study this law, we will learn something about nature and also find that such a law limits the efficiency of heat engines. We will find that heat engines with the greatest possible theoretical efficiency would have to use reversible processes, and even they cannot convert all heat transfer into doing work. [\[link\]](#) summarizes the simpler thermodynamic processes and their definitions.

Isobaric	Constant pressure $W = P\Delta V$
Isochoric	Constant volume $W = 0$
Isothermal	Constant temperature $Q = W$
Adiabatic	No heat transfer $Q = 0$

Summary of Simple Thermodynamic Processes

Note:

PhET Explorations: States of Matter

Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.

https://phet.colorado.edu/sims/html/states-of-matter/latest/states-of-matter_en.html

Section Summary

- One of the important implications of the first law of thermodynamics is that machines can be harnessed to do work that humans previously

did by hand or by external energy supplies such as running water or the heat of the Sun. A machine that uses heat transfer to do work is known as a heat engine.

- There are several simple processes, used by heat engines, that flow from the first law of thermodynamics. Among them are the isobaric, isochoric, isothermal and adiabatic processes.
- These processes differ from one another based on how they affect pressure, volume, temperature, and heat transfer.
- If the work done is performed on the outside environment, work (W) will be a positive value. If the work done is done to the heat engine system, work (W) will be a negative value.
- Some thermodynamic processes, including isothermal and adiabatic processes, are reversible in theory; that is, both the thermodynamic system and the environment can be returned to their initial states. However, because of loss of energy owing to the second law of thermodynamics, complete reversibility does not work in practice.

Conceptual Questions

Exercise:

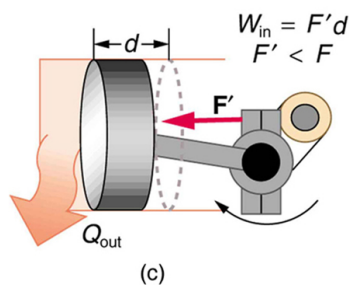
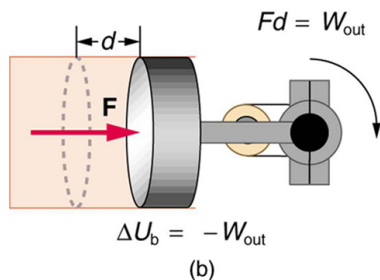
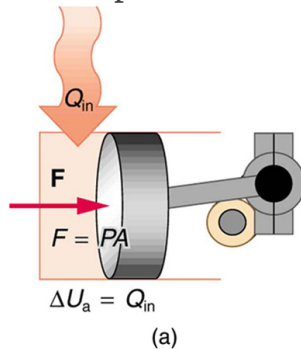
Problem:

A great deal of effort, time, and money has been spent in the quest for the so-called perpetual-motion machine, which is defined as a hypothetical machine that operates or produces useful work indefinitely and/or a hypothetical machine that produces more work or energy than it consumes. Explain, in terms of heat engines and the first law of thermodynamics, why or why not such a machine is likely to be constructed.

Exercise:

Problem:

One method of converting heat transfer into doing work is for heat transfer into a gas to take place, which expands, doing work on a piston, as shown in the figure below. (a) Is the heat transfer converted directly to work in an isobaric process, or does it go through another form first? Explain your answer. (b) What about in an isothermal process? (c) What about in an adiabatic process (where heat transfer occurred prior to the adiabatic process)?



Exercise:

Problem:

Would the previous question make any sense for an isochoric process? Explain your answer.

Exercise:**Problem:**

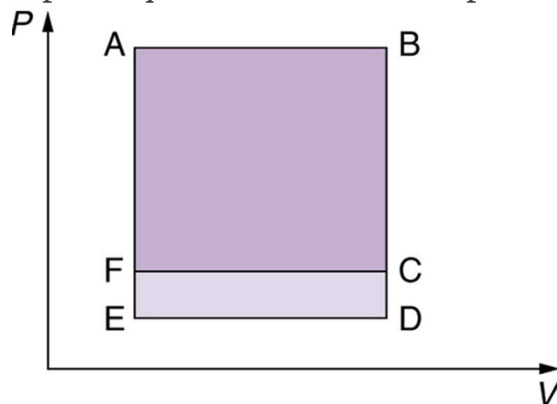
We ordinarily say that $\Delta U = 0$ for an isothermal process. Does this assume no phase change takes place? Explain your answer.

Exercise:**Problem:**

The temperature of a rapidly expanding gas decreases. Explain why in terms of the first law of thermodynamics. (Hint: Consider whether the gas does work and whether heat transfer occurs rapidly into the gas through conduction.)

Exercise:**Problem:**

Which cyclical process represented by the two closed loops, ABCFA and ABDEA, on the PV diagram in the figure below produces the greatest *net* work? Is that process also the one with the smallest work input required to return it to point A? Explain your responses.



The two cyclical processes shown on this PV diagram start with and return the system to the conditions at point A, but they follow

different paths and produce different amounts of work.

Exercise:

Problem:

A real process may be nearly adiabatic if it occurs over a very short time. How does the short time span help the process to be adiabatic?

Exercise:

Problem:

It is unlikely that a process can be isothermal unless it is a very slow process. Explain why. Is the same true for isobaric and isochoric processes? Explain your answer.

Problem Exercises

Exercise:

Problem:

A car tire contains 0.0380 m^3 of air at a pressure of $2.20 \times 10^5 \text{ N/m}^2$ (about 32 psi). How much more internal energy does this gas have than the same volume has at zero gauge pressure (which is equivalent to normal atmospheric pressure)?

Solution:

$$6.77 \times 10^3 \text{ J}$$

Exercise:

Problem:

A helium-filled toy balloon has a gauge pressure of 0.200 atm and a volume of 10.0 L. How much greater is the internal energy of the helium in the balloon than it would be at zero gauge pressure?

Exercise:**Problem:**

Steam to drive an old-fashioned steam locomotive is supplied at a constant gauge pressure of $1.75 \times 10^6 \text{ N/m}^2$ (about 250 psi) to a piston with a 0.200-m radius. (a) By calculating $P\Delta V$, find the work done by the steam when the piston moves 0.800 m. Note that this is the net work output, since gauge pressure is used. (b) Now find the amount of work by calculating the force exerted times the distance traveled. Is the answer the same as in part (a)?

Solution:

(a) $W = P\Delta V = 1.76 \times 10^5 \text{ J}$

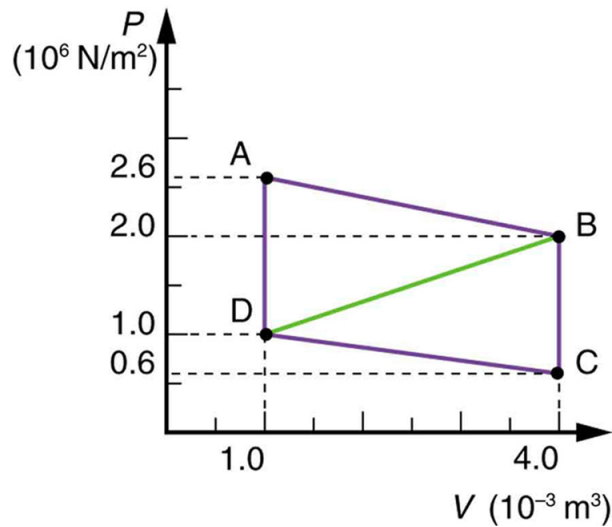
(b) $W = Fd = 1.76 \times 10^5 \text{ J}$. Yes, the answer is the same.

Exercise:**Problem:**

A hand-driven tire pump has a piston with a 2.50-cm diameter and a maximum stroke of 30.0 cm. (a) How much work do you do in one stroke if the average gauge pressure is $2.40 \times 10^5 \text{ N/m}^2$ (about 35 psi)? (b) What average force do you exert on the piston, neglecting friction and gravitational force?

Exercise:**Problem:**

Calculate the net work output of a heat engine following path ABCDA in the figure below.



Solution:

$$W = 4.5 \times 10^3 \text{ J}$$

Exercise:

Problem:

What is the net work output of a heat engine that follows path ABDA in the figure above, with a straight line from B to D? Why is the work output less than for path ABCDA? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Exercise:

Problem: Unreasonable Results

What is wrong with the claim that a cyclical heat engine does 4.00 kJ of work on an input of 24.0 kJ of heat transfer while 16.0 kJ of heat transfers to the environment?

Solution:

W is not equal to the difference between the heat input and the heat output.

Exercise:

Problem:

- (a) A cyclical heat engine, operating between temperatures of 450°C and 150°C produces 4.00 MJ of work on a heat transfer of 5.00 MJ into the engine. How much heat transfer occurs to the environment?
(b) What is unreasonable about the engine? (c) Which premise is unreasonable?

Exercise:**Problem: Construct Your Own Problem**

Consider a car's gasoline engine. Construct a problem in which you calculate the maximum efficiency this engine can have. Among the things to consider are the effective hot and cold reservoir temperatures. Compare your calculated efficiency with the actual efficiency of car engines.

Exercise:**Problem: Construct Your Own Problem**

Consider a car trip into the mountains. Construct a problem in which you calculate the overall efficiency of the car for the trip as a ratio of kinetic and potential energy gained to fuel consumed. Compare this efficiency to the thermodynamic efficiency quoted for gasoline engines and discuss why the thermodynamic efficiency is so much greater. Among the factors to be considered are the gain in altitude and speed, the mass of the car, the distance traveled, and typical fuel economy.

Glossary

heat engine

a machine that uses heat transfer to do work

isobaric process

constant-pressure process in which a gas does work

isochoric process

a constant-volume process

isothermal process

a constant-temperature process

adiabatic process

a process in which no heat transfer takes place

reversible process

a process in which both the heat engine system and the external environment theoretically can be returned to their original states

Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

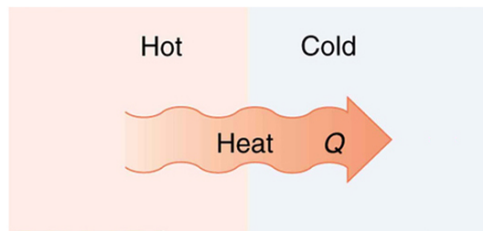
- State the expressions of the second law of thermodynamics.
- Calculate the efficiency and carbon dioxide emission of a coal-fired electricity plant, using second law characteristics.
- Describe and define the Otto cycle.



These ice floes melt during the Arctic summer. Some of them refreeze in the winter, but the second law of thermodynamics predicts that it would be extremely unlikely for the water molecules contained in these particular floes to reform the distinctive alligator-like shape they formed when the picture was taken in the summer of 2009. (credit: Patrick Kelley, U.S. Coast Guard, U.S. Geological Survey)

The second law of thermodynamics deals with the direction taken by spontaneous processes. Many processes occur spontaneously in one direction only—that is, they are irreversible, under a given set of

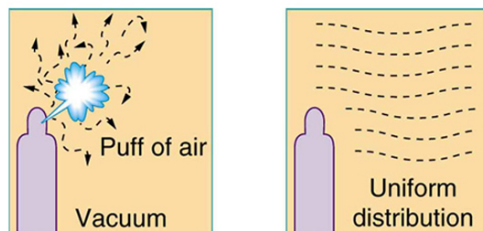
conditions. Although irreversibility is seen in day-to-day life—a broken glass does not resume its original state, for instance—complete irreversibility is a statistical statement that cannot be seen during the lifetime of the universe. More precisely, an **irreversible process** is one that depends on path. If the process can go in only one direction, then the reverse path differs fundamentally and the process cannot be reversible. For example, as noted in the previous section, heat involves the transfer of energy from higher to lower temperature. A cold object in contact with a hot one never gets colder, transferring heat to the hot object and making it hotter. Furthermore, mechanical energy, such as kinetic energy, can be completely converted to thermal energy by friction, but the reverse is impossible. A hot stationary object never spontaneously cools off and starts moving. Yet another example is the expansion of a puff of gas introduced into one corner of a vacuum chamber. The gas expands to fill the chamber, but it never regroups in the corner. The random motion of the gas molecules could take them all back to the corner, but this is never observed to happen. (See [\[link\]](#).)



(a)



(b)



(c)

Examples of one-way processes in nature.

(a) Heat transfer occurs spontaneously from hot to cold and not from cold to hot. (b) The brakes of this car convert its kinetic energy to heat transfer to the environment. The reverse process is impossible. (c) The burst of gas let into this vacuum chamber quickly expands to uniformly fill every part of the chamber. The random motions of the gas molecules will never return them to the corner.

The fact that certain processes never occur suggests that there is a law forbidding them to occur. The first law of thermodynamics would allow them to occur—none of those processes violate conservation of energy. The law that forbids these processes is called the second law of thermodynamics. We shall see that the second law can be stated in many ways that may seem different, but which in fact are equivalent. Like all natural laws, the second law of thermodynamics gives insights into nature, and its several statements imply that it is broadly applicable, fundamentally affecting many apparently disparate processes.

The already familiar direction of heat transfer from hot to cold is the basis of our first version of the **second law of thermodynamics**.

Note:

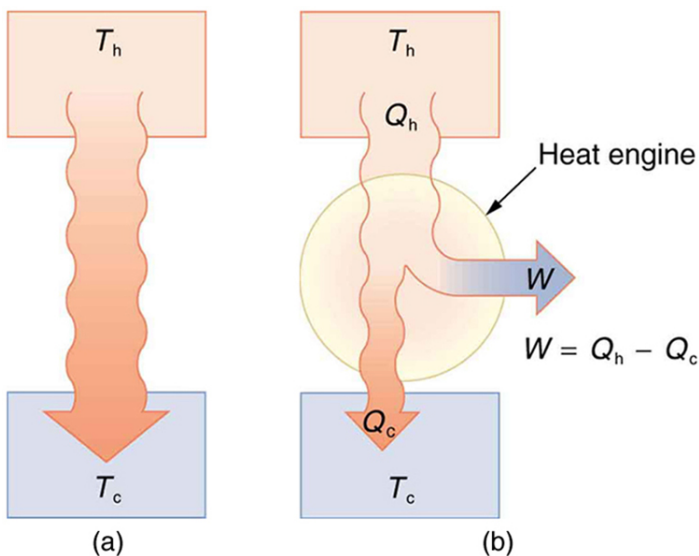
The Second Law of Thermodynamics (first expression)

Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction.

Another way of stating this: It is impossible for any process to have as its sole result heat transfer from a cooler to a hotter object.

Heat Engines

Now let us consider a device that uses heat transfer to do work. As noted in the previous section, such a device is called a heat engine, and one is shown schematically in [\[link\]](#)(b). Gasoline and diesel engines, jet engines, and steam turbines are all heat engines that do work by using part of the heat transfer from some source. Heat transfer from the hot object (or hot reservoir) is denoted as Q_h , while heat transfer into the cold object (or cold reservoir) is Q_c , and the work done by the engine is W . The temperatures of the hot and cold reservoirs are T_h and T_c , respectively.



- (a) Heat transfer occurs spontaneously from a hot object to a cold one, consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the heat transfer to do work. The hot and cold objects are called the hot and cold reservoirs. Q_h is the heat transfer out of the hot reservoir, W is the work output, and Q_c is the heat transfer into the cold reservoir.

Because the hot reservoir is heated externally, which is energy intensive, it is important that the work is done as efficiently as possible. In fact, we would like W to equal Q_h , and for there to be no heat transfer to the environment ($Q_c = 0$). Unfortunately, this is impossible. The **second law of thermodynamics** also states, with regard to using heat transfer to do work (the second expression of the second law):

Note:

The Second Law of Thermodynamics (second expression)

It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.

A **cyclical process** brings a system, such as the gas in a cylinder, back to its original state at the end of every cycle. Most heat engines, such as reciprocating piston engines and rotating turbines, use cyclical processes. The second law, just stated in its second form, clearly states that such engines cannot have perfect conversion of heat transfer into work done. Before going into the underlying reasons for the limits on converting heat transfer into work, we need to explore the relationships among W , Q_h , and Q_c , and to define the efficiency of a cyclical heat engine. As noted, a cyclical process brings the system back to its original condition at the end of every cycle. Such a system's internal energy U is the same at the beginning and end of every cycle—that is, $\Delta U = 0$. The first law of thermodynamics states that

Equation:

$$\Delta U = Q - W,$$

where Q is the *net* heat transfer during the cycle ($Q = Q_h - Q_c$) and W is the net work done by the system. Since $\Delta U = 0$ for a complete cycle, we

have

Equation:

$$0 = Q - W,$$

so that

Equation:

$$W = Q.$$

Thus the net work done by the system equals the net heat transfer into the system, or

Equation:

$$W = Q_h - Q_c \text{ (cyclical process),}$$

just as shown schematically in [\[link\]](#)(b). The problem is that in all processes, there is some heat transfer Q_c to the environment—and usually a very significant amount at that.

In the conversion of energy to work, we are always faced with the problem of getting less out than we put in. We define *conversion efficiency* Eff to be the ratio of useful work output to the energy input (or, in other words, the ratio of what we get to what we spend). In that spirit, we define the efficiency of a heat engine to be its net work output W divided by heat transfer to the engine Q_h ; that is,

Equation:

$$\text{Eff} = \frac{W}{Q_h}.$$

Since $W = Q_h - Q_c$ in a cyclical process, we can also express this as

Equation:

$$\text{Eff} = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h} \text{ (cyclical process),}$$

making it clear that an efficiency of 1, or 100%, is possible only if there is no heat transfer to the environment ($Q_c = 0$). Note that all Q s are positive. The direction of heat transfer is indicated by a plus or minus sign. For example, Q_c is out of the system and so is preceded by a minus sign.

Example:

Daily Work Done by a Coal-Fired Power Station, Its Efficiency and Carbon Dioxide Emissions

A coal-fired power station is a huge heat engine. It uses heat transfer from burning coal to do work to turn turbines, which are used to generate electricity. In a single day, a large coal power station has 2.50×10^{14} J of heat transfer from coal and 1.48×10^{14} J of heat transfer into the environment. (a) What is the work done by the power station? (b) What is the efficiency of the power station? (c) In the combustion process, the following chemical reaction occurs: $C + O_2 \rightarrow CO_2$. This implies that every 12 kg of coal puts $12 \text{ kg} + 16 \text{ kg} + 16 \text{ kg} = 44 \text{ kg}$ of carbon dioxide into the atmosphere. Assuming that 1 kg of coal can provide 2.5×10^6 J of heat transfer upon combustion, how much CO_2 is emitted per day by this power plant?

Strategy for (a)

We can use $W = Q_h - Q_c$ to find the work output W , assuming a cyclical process is used in the power station. In this process, water is boiled under pressure to form high-temperature steam, which is used to run steam turbine-generators, and then condensed back to water to start the cycle again.

Solution for (a)

Work output is given by:

Equation:

$$W = Q_h - Q_c.$$

Substituting the given values:

Equation:

$$\begin{aligned} W &= 2.50 \times 10^{14} \text{ J} - 1.48 \times 10^{14} \text{ J} \\ &= 1.02 \times 10^{14} \text{ J}. \end{aligned}$$

Strategy for (b)

The efficiency can be calculated with $\text{Eff} = \frac{W}{Q_h}$ since Q_h is given and work W was found in the first part of this example.

Solution for (b)

Efficiency is given by: $\text{Eff} = \frac{W}{Q_h}$. The work W was just found to be $1.02 \times 10^{14} \text{ J}$, and Q_h is given, so the efficiency is

Equation:

$$\begin{aligned} \text{Eff} &= \frac{1.02 \times 10^{14} \text{ J}}{2.50 \times 10^{14} \text{ J}} \\ &= 0.408, \text{ or } 40.8\% \end{aligned}$$

Strategy for (c)

The daily consumption of coal is calculated using the information that each day there is $2.50 \times 10^{14} \text{ J}$ of heat transfer from coal. In the combustion process, we have $\text{C} + \text{O}_2 \rightarrow \text{CO}_2$. So every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of CO_2 into the atmosphere.

Solution for (c)

The daily coal consumption is

Equation:

$$\frac{2.50 \times 10^{14} \text{ J}}{2.50 \times 10^6 \text{ J/kg}} = 1.0 \times 10^8 \text{ kg}.$$

Assuming that the coal is pure and that all the coal goes toward producing carbon dioxide, the carbon dioxide produced per day is

Equation:

$$1.0 \times 10^8 \text{ kg coal} \times \frac{44 \text{ kg CO}_2}{12 \text{ kg coal}} = 3.7 \times 10^8 \text{ kg CO}_2.$$

This is 370,000 metric tons of CO_2 produced every day.

Discussion

If all the work output is converted to electricity in a period of one day, the average power output is 1180 MW (this is left to you as an end-of-chapter problem). This value is about the size of a large-scale conventional power plant. The efficiency found is acceptably close to the value of 42% given for coal power stations. It means that fully 59.2% of the energy is heat transfer to the environment, which usually results in warming lakes, rivers, or the ocean near the power station, and is implicated in a warming planet generally. While the laws of thermodynamics limit the efficiency of such plants—including plants fired by nuclear fuel, oil, and natural gas—the heat transfer to the environment could be, and sometimes is, used for heating homes or for industrial processes. The generally low cost of energy has not made it economical to make better use of the waste heat transfer from most heat engines. Coal-fired power plants produce the greatest amount of CO_2 per unit energy output (compared to natural gas or oil), making coal the least efficient fossil fuel.

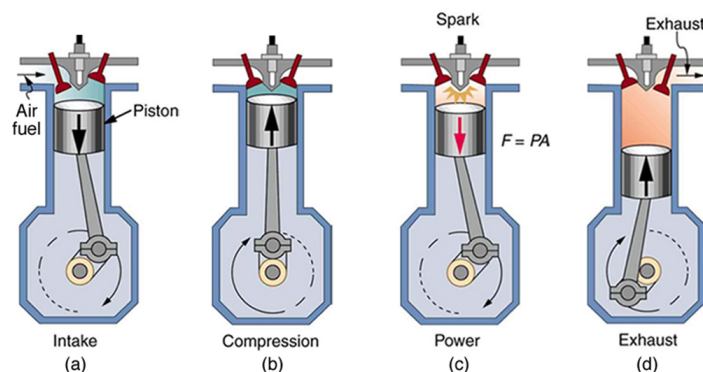
With the information given in [\[link\]](#), we can find characteristics such as the efficiency of a heat engine without any knowledge of how the heat engine operates, but looking further into the mechanism of the engine will give us greater insight. [\[link\]](#) illustrates the operation of the common four-stroke gasoline engine. The four steps shown complete this heat engine's cycle, bringing the gasoline-air mixture back to its original condition.

The **Otto cycle** shown in [\[link\]](#)(a) is used in four-stroke internal combustion engines, although in fact the true Otto cycle paths do not correspond exactly to the strokes of the engine.

The adiabatic process AB corresponds to the nearly adiabatic compression stroke of the gasoline engine. In both cases, work is done on the system (the gas mixture in the cylinder), increasing its temperature and pressure. Along path BC of the Otto cycle, heat transfer Q_h into the gas occurs at constant volume, causing a further increase in pressure and temperature. This process corresponds to burning fuel in an internal combustion engine, and takes place so rapidly that the volume is nearly constant. Path CD in the Otto cycle is an adiabatic expansion that does work on the outside world,

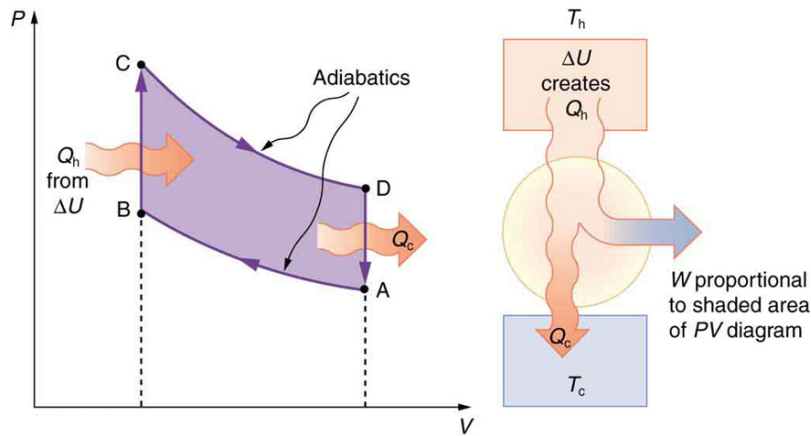
just as the power stroke of an internal combustion engine does in its nearly adiabatic expansion. The work done by the system along path CD is greater than the work done on the system along path AB, because the pressure is greater, and so there is a net work output. Along path DA in the Otto cycle, heat transfer Q_c from the gas at constant volume reduces its temperature and pressure, returning it to its original state. In an internal combustion engine, this process corresponds to the exhaust of hot gases and the intake of an air-gasoline mixture at a considerably lower temperature. In both cases, heat transfer into the environment occurs along this final path.

The net work done by a cyclical process is the area inside the closed path on a PV diagram, such as that inside path ABCDA in [\[link\]](#). Note that in every imaginable cyclical process, it is absolutely necessary for heat transfer from the system to occur in order to get a net work output. In the Otto cycle, heat transfer occurs along path DA. If no heat transfer occurs, then the return path is the same, and the net work output is zero. The lower the temperature on the path AB, the less work has to be done to compress the gas. The area inside the closed path is then greater, and so the engine does more work and is thus more efficient. Similarly, the higher the temperature along path CD, the more work output there is. (See [\[link\]](#).) So efficiency is related to the temperatures of the hot and cold reservoirs. In the next section, we shall see what the absolute limit to the efficiency of a heat engine is, and how it is related to temperature.



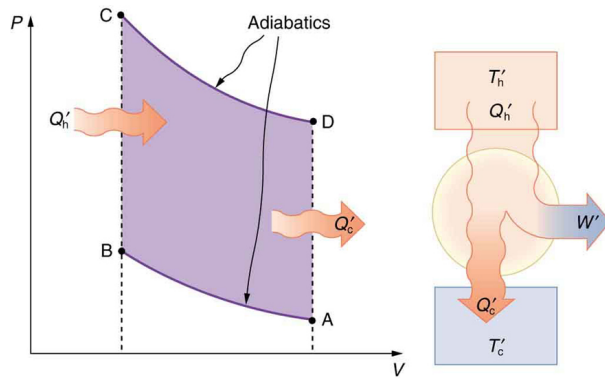
In the four-stroke internal combustion gasoline engine, heat transfer into

work takes place in the cyclical process shown here. The piston is connected to a rotating crankshaft, which both takes work out of and does work on the gas in the cylinder. (a) Air is mixed with fuel during the intake stroke. (b) During the compression stroke, the air-fuel mixture is rapidly compressed in a nearly adiabatic process, as the piston rises with the valves closed. Work is done on the gas. (c) The power stroke has two distinct parts. First, the air-fuel mixture is ignited, converting chemical potential energy into thermal energy almost instantaneously, which leads to a great increase in pressure. Then the piston descends, and the gas does work by exerting a force through a distance in a nearly adiabatic process. (d) The exhaust stroke expels the hot gas to prepare the engine for another cycle, starting again with the intake stroke.



PV diagram for a simplified Otto cycle, analogous to that employed in an internal combustion engine. Point A corresponds to the start of the compression stroke of an internal combustion engine. Paths AB and CD are adiabatic and correspond to the compression and power strokes of an internal combustion engine, respectively.

Paths BC and DA are isochoric and accomplish similar results to the ignition and exhaust-intake portions, respectively, of the internal combustion engine's cycle. Work is done on the gas along path AB, but more work is done by the gas along path CD, so that there is a net work output.



This Otto cycle produces a greater work output than the one in [\[link\]](#), because the starting temperature of path CD is higher and the starting temperature of path AB is lower. The area inside the loop is greater, corresponding to greater net work output.

Section Summary

- The two expressions of the second law of thermodynamics are: (i) Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction; and (ii) It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.
- Irreversible processes depend on path and do not return to their original state. Cyclical processes are processes that return to their original state at the end of every cycle.
- In a cyclical process, such as a heat engine, the net work done by the system equals the net heat transfer into the system, or $W = Q_h - Q_c$, where Q_h is the heat transfer from the hot object (hot reservoir), and Q_c is the heat transfer into the cold object (cold reservoir).

- Efficiency can be expressed as $\text{Eff} = \frac{W}{Q_h}$, the ratio of work output divided by the amount of energy input.
- The four-stroke gasoline engine is often explained in terms of the Otto cycle, which is a repeating sequence of processes that convert heat into work.

Conceptual Questions

Exercise:

Problem:

Imagine you are driving a car up Pike's Peak in Colorado. To raise a car weighing 1000 kilograms a distance of 100 meters would require about a million joules. You could raise a car 12.5 kilometers with the energy in a gallon of gas. Driving up Pike's Peak (a mere 3000-meter climb) should consume a little less than a quart of gas. But other considerations have to be taken into account. Explain, in terms of efficiency, what factors may keep you from realizing your ideal energy use on this trip.

Exercise:

Problem:

Is a temperature difference necessary to operate a heat engine? State why or why not.

Exercise:

Problem:

Definitions of efficiency vary depending on how energy is being converted. Compare the definitions of efficiency for the human body and heat engines. How does the definition of efficiency in each relate to the type of energy being converted into doing work?

Exercise:

Problem:

Why—other than the fact that the second law of thermodynamics says reversible engines are the most efficient—should heat engines employing reversible processes be more efficient than those employing irreversible processes? Consider that dissipative mechanisms are one cause of irreversibility.

Problem Exercises**Exercise:****Problem:**

A certain heat engine does 10.0 kJ of work and 8.50 kJ of heat transfer occurs to the environment in a cyclical process. (a) What was the heat transfer into this engine? (b) What was the engine's efficiency?

Solution:

(a) 18.5 kJ

(b) 54.1%

Exercise:**Problem:**

With 2.56×10^6 J of heat transfer into this engine, a given cyclical heat engine can do only 1.50×10^5 J of work. (a) What is the engine's efficiency? (b) How much heat transfer to the environment takes place?

Exercise:

Problem:

(a) What is the work output of a cyclical heat engine having a 22.0% efficiency and 6.00×10^9 J of heat transfer into the engine? (b) How much heat transfer occurs to the environment?

Solution:

(a) 1.32×10^9 J

(b) 4.68×10^9 J

Exercise:**Problem:**

(a) What is the efficiency of a cyclical heat engine in which 75.0 kJ of heat transfer occurs to the environment for every 95.0 kJ of heat transfer into the engine? (b) How much work does it produce for 100 kJ of heat transfer into the engine?

Exercise:**Problem:**

The engine of a large ship does 2.00×10^8 J of work with an efficiency of 5.00%. (a) How much heat transfer occurs to the environment? (b) How many barrels of fuel are consumed, if each barrel produces 6.00×10^9 J of heat transfer when burned?

Solution:

(a) 3.80×10^9 J

(b) 0.667 barrels

Exercise:

Problem:

(a) How much heat transfer occurs to the environment by an electrical power station that uses 1.25×10^{14} J of heat transfer into the engine with an efficiency of 42.0%? (b) What is the ratio of heat transfer to the environment to work output? (c) How much work is done?

Exercise:**Problem:**

Assume that the turbines at a coal-powered power plant were upgraded, resulting in an improvement in efficiency of 3.32%. Assume that prior to the upgrade the power station had an efficiency of 36% and that the heat transfer into the engine in one day is still the same at 2.50×10^{14} J. (a) How much more electrical energy is produced due to the upgrade? (b) How much less heat transfer occurs to the environment due to the upgrade?

Solution:

(a) 8.30×10^{12} J, which is 3.32% of 2.50×10^{14} J .

(b) -8.30×10^{12} J, where the negative sign indicates a reduction in heat transfer to the environment.

Exercise:

Problem:

This problem compares the energy output and heat transfer to the environment by two different types of nuclear power stations—one with the normal efficiency of 34.0%, and another with an improved efficiency of 40.0%. Suppose both have the same heat transfer into the engine in one day, 2.50×10^{14} J. (a) How much more electrical energy is produced by the more efficient power station? (b) How much less heat transfer occurs to the environment by the more efficient power station? (One type of more efficient nuclear power station, the gas-cooled reactor, has not been reliable enough to be economically feasible in spite of its greater efficiency.)

Glossary

irreversible process

any process that depends on path direction

second law of thermodynamics

heat transfer flows from a hotter to a cooler object, never the reverse, and some heat energy in any process is lost to available work in a cyclical process

cyclical process

a process in which the path returns to its original state at the end of every cycle

Otto cycle

a thermodynamic cycle, consisting of a pair of adiabatic processes and a pair of isochoric processes, that converts heat into work, e.g., the four-stroke engine cycle of intake, compression, ignition, and exhaust

Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated

- Identify a Carnot cycle.
- Calculate maximum theoretical efficiency of a nuclear reactor.
- Explain how dissipative processes affect the ideal Carnot engine.



This novelty toy, known as the drinking bird, is an example of Carnot's engine. It contains methylene chloride (mixed with a dye) in the abdomen, which boils at a very low temperature—about 100°F. To operate, one gets the bird's head wet. As the water evaporates, fluid moves up into the head, causing the bird to become top-heavy and dip forward back into the water. This cools down the methylene chloride in the head, and it moves back into the abdomen, causing the bird to become bottom heavy and tip up. Except for a very small input of energy—the original head-wetting—the bird becomes a perpetual motion machine of sorts. (credit: Arabesk.nl, Wikimedia Commons)

We know from the second law of thermodynamics that a heat engine cannot be 100% efficient, since there must always be some heat transfer Q_c to the environment, which is often called waste heat. How efficient, then, can a heat engine be? This question was answered at a theoretical level in 1824 by a young French engineer, Sadi Carnot (1796–1832), in his study of the then-emerging heat engine technology crucial to the Industrial Revolution. He devised a theoretical cycle, now called the **Carnot cycle**, which is the most efficient cyclical process possible. The second law of thermodynamics can be restated in terms of the Carnot cycle, and so what Carnot actually discovered was this fundamental law. Any heat engine employing the Carnot cycle is called a **Carnot engine**.

What is crucial to the Carnot cycle—and, in fact, defines it—is that only reversible processes are used. Irreversible processes involve dissipative factors, such as friction and turbulence. This increases heat transfer Q_c to the environment and reduces the efficiency of the engine. Obviously, then, reversible processes are superior.

Note:

Carnot Engine

Stated in terms of reversible processes, the **second law of thermodynamics** has a third form:

A Carnot engine operating between two given temperatures has the greatest possible efficiency of any heat engine operating between these two temperatures. Furthermore, all engines employing only reversible processes have this same maximum efficiency when operating between the same given temperatures.

[\[link\]](#) shows the PV diagram for a Carnot cycle. The cycle comprises two isothermal and two adiabatic processes. Recall that both isothermal and adiabatic processes are, in principle, reversible.

Carnot also determined the efficiency of a perfect heat engine—that is, a Carnot engine. It is always true that the efficiency of a cyclical heat engine is given by:

Equation:

$$\text{Eff} = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h}.$$

What Carnot found was that for a perfect heat engine, the ratio Q_c/Q_h equals the ratio of the absolute temperatures of the heat reservoirs. That is, $Q_c/Q_h = T_c/T_h$ for a Carnot engine, so that the maximum or **Carnot efficiency** Eff_C is given by

Equation:

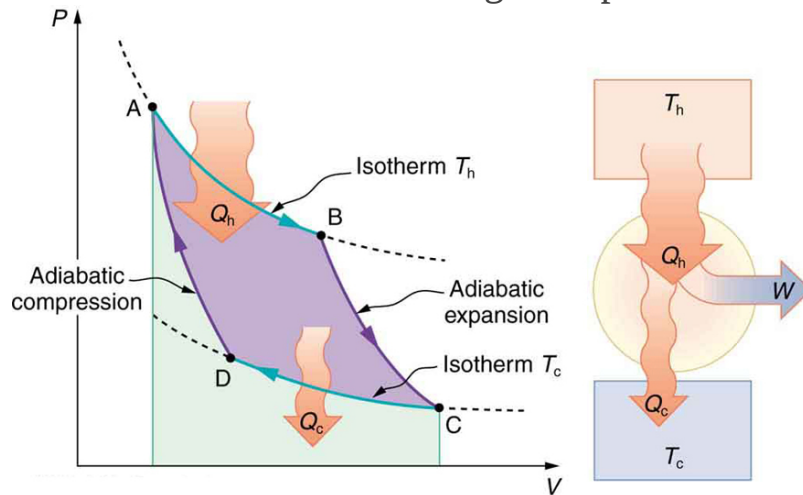
$$\text{Eff}_C = 1 - \frac{T_c}{T_h},$$

where T_h and T_c are in kelvins (or any other absolute temperature scale). No real heat engine can do as well as the Carnot efficiency—an actual efficiency of about 0.7 of this maximum is usually the best that can be accomplished. But the ideal Carnot engine, like the drinking bird above, while a fascinating novelty, has zero power. This makes it unrealistic for any applications.

Carnot's interesting result implies that 100% efficiency would be possible only if $T_c = 0$ K—that is, only if the cold reservoir were at absolute zero, a practical and theoretical impossibility. But the physical implication is this—the only way to have all heat transfer go into doing work is to remove *all* thermal energy, and this requires a cold reservoir at absolute zero.

It is also apparent that the greatest efficiencies are obtained when the ratio T_c/T_h is as small as possible. Just as discussed for the Otto cycle in the previous section, this means that efficiency is greatest for the highest possible temperature of the hot reservoir and lowest possible temperature of the cold reservoir. (This setup increases the area inside the closed loop on the PV diagram; also, it seems reasonable that the greater the temperature

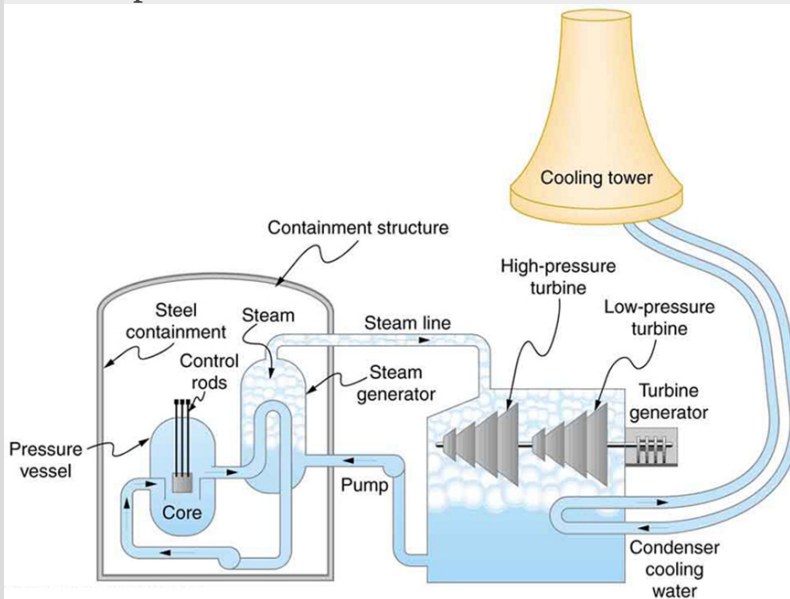
difference, the easier it is to divert the heat transfer to work.) The actual reservoir temperatures of a heat engine are usually related to the type of heat source and the temperature of the environment into which heat transfer occurs. Consider the following example.



PV diagram for a Carnot cycle, employing only reversible isothermal and adiabatic processes. Heat transfer Q_h occurs into the working substance during the isothermal path AB, which takes place at constant temperature T_h . Heat transfer Q_c occurs out of the working substance during the isothermal path CD, which takes place at constant temperature T_c . The net work output W equals the area inside the path ABCDA. Also shown is a schematic of a Carnot engine operating between hot and cold reservoirs at temperatures T_h and T_c . Any heat engine using reversible processes and operating between these two temperatures will have the same maximum efficiency as the Carnot engine.

Example:**Maximum Theoretical Efficiency for a Nuclear Reactor**

A nuclear power reactor has pressurized water at 300°C . (Higher temperatures are theoretically possible but practically not, due to limitations with materials used in the reactor.) Heat transfer from this water is a complex process (see [\[link\]](#)). Steam, produced in the steam generator, is used to drive the turbine-generators. Eventually the steam is condensed to water at 27°C and then heated again to start the cycle over. Calculate the maximum theoretical efficiency for a heat engine operating between these two temperatures.



Schematic diagram of a pressurized water nuclear reactor and the steam turbines that convert work into electrical energy. Heat exchange is used to generate steam, in part to avoid contamination of the generators with radioactivity. Two turbines are used because this is less expensive than operating a single generator that produces the same amount of electrical energy. The steam is condensed to liquid before being returned to the heat exchanger, to keep exit steam pressure low and aid the flow of steam through the turbines (equivalent to using a

lower-temperature cold reservoir). The considerable energy associated with condensation must be dissipated into the local environment; in this example, a cooling tower is used so there is no direct heat transfer to an aquatic environment. (Note that the water going to the cooling tower does not come into contact with the steam flowing over the turbines.)

Strategy

Since temperatures are given for the hot and cold reservoirs of this heat engine, $Eff_C = 1 - \frac{T_c}{T_h}$ can be used to calculate the Carnot (maximum theoretical) efficiency. Those temperatures must first be converted to kelvins.

Solution

The hot and cold reservoir temperatures are given as 300°C and 27.0°C , respectively. In kelvins, then, $T_h = 573\text{ K}$ and $T_c = 300\text{ K}$, so that the maximum efficiency is

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

Thus,

Equation:

$$\begin{aligned} Eff_C &= 1 - \frac{300\text{ K}}{573\text{ K}} \\ &= 0.476, \text{ or } 47.6\%. \end{aligned}$$

Discussion

A typical nuclear power station's actual efficiency is about 35%, a little better than 0.7 times the maximum possible value, a tribute to superior engineering. Electrical power stations fired by coal, oil, and natural gas have greater actual efficiencies (about 42%), because their boilers can reach higher temperatures and pressures. The cold reservoir temperature in

any of these power stations is limited by the local environment. [\[link\]](#) shows (a) the exterior of a nuclear power station and (b) the exterior of a coal-fired power station. Both have cooling towers into which water from the condenser enters the tower near the top and is sprayed downward, cooled by evaporation.



(a)



(b)

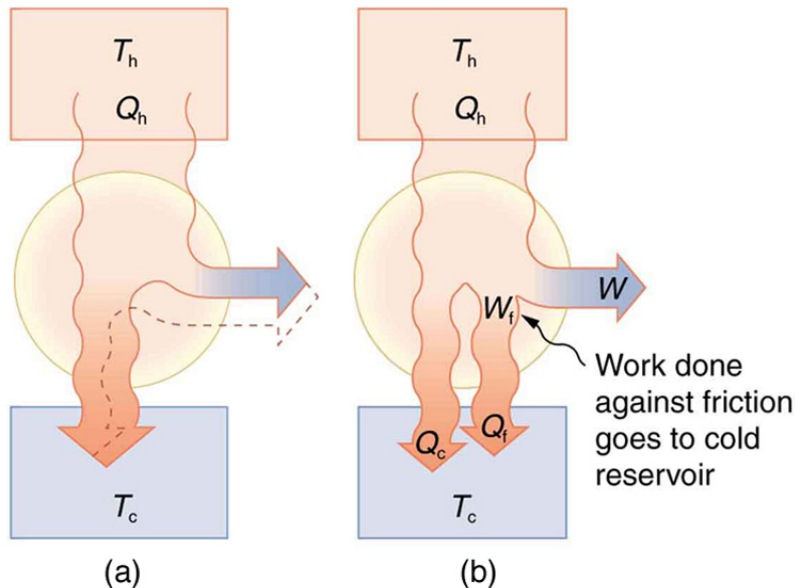
(a) A nuclear power station (credit: BlatantWorld.com) and (b) a coal-fired power station.

Both have cooling towers in which water evaporates into the environment, representing Q_c .

The nuclear reactor, which supplies Q_h , is housed inside

the dome-shaped containment buildings. (credit: Robert & Mihaela Vicol, publicphoto.org)

Since all real processes are irreversible, the actual efficiency of a heat engine can never be as great as that of a Carnot engine, as illustrated in [\[link\]](#)(a). Even with the best heat engine possible, there are always dissipative processes in peripheral equipment, such as electrical transformers or car transmissions. These further reduce the overall efficiency by converting some of the engine's work output back into heat transfer, as shown in [\[link\]](#)(b).



Real heat engines are less efficient than Carnot engines. (a) Real engines use irreversible processes, reducing the heat transfer to work. Solid lines represent the actual process; the dashed lines are what a Carnot engine would do between the same two reservoirs. (b) Friction and other dissipative processes in the output mechanisms of a heat engine convert some

of its work output into heat transfer to the environment.

Section Summary

- The Carnot cycle is a theoretical cycle that is the most efficient cyclical process possible. Any engine using the Carnot cycle, which uses only reversible processes (adiabatic and isothermal), is known as a Carnot engine.
- Any engine that uses the Carnot cycle enjoys the maximum theoretical efficiency.
- While Carnot engines are ideal engines, in reality, no engine achieves Carnot's theoretical maximum efficiency, since dissipative processes, such as friction, play a role. Carnot cycles without heat loss may be possible at absolute zero, but this has never been seen in nature.

Conceptual Questions

Exercise:

Problem:

Think about the drinking bird at the beginning of this section ([link](#)). Although the bird enjoys the theoretical maximum efficiency possible, if left to its own devices over time, the bird will cease “drinking.” What are some of the dissipative processes that might cause the bird's motion to cease?

Exercise:

Problem:

Can improved engineering and materials be employed in heat engines to reduce heat transfer into the environment? Can they eliminate heat transfer into the environment entirely?

Exercise:

Problem:

Does the second law of thermodynamics alter the conservation of energy principle?

Problem Exercises**Exercise:****Problem:**

A certain gasoline engine has an efficiency of 30.0%. What would the hot reservoir temperature be for a Carnot engine having that efficiency, if it operates with a cold reservoir temperature of 200°C?

Solution:

403°C

Exercise:**Problem:**

A gas-cooled nuclear reactor operates between hot and cold reservoir temperatures of 700°C and 27.0°C. (a) What is the maximum efficiency of a heat engine operating between these temperatures? (b) Find the ratio of this efficiency to the Carnot efficiency of a standard nuclear reactor (found in [\[link\]](#)).

Exercise:**Problem:**

(a) What is the hot reservoir temperature of a Carnot engine that has an efficiency of 42.0% and a cold reservoir temperature of 27.0°C? (b) What must the hot reservoir temperature be for a real heat engine that achieves 0.700 of the maximum efficiency, but still has an efficiency of 42.0% (and a cold reservoir at 27.0°C)? (c) Does your answer imply practical limits to the efficiency of car gasoline engines?

Solution:

(a) 244°C

(b) 477°C

(c) Yes, since automobiles engines cannot get too hot without overheating, their efficiency is limited.

Exercise:**Problem:**

Steam locomotives have an efficiency of 17.0% and operate with a hot steam temperature of 425°C. (a) What would the cold reservoir temperature be if this were a Carnot engine? (b) What would the maximum efficiency of this steam engine be if its cold reservoir temperature were 150°C?

Exercise:**Problem:**

Practical steam engines utilize 450°C steam, which is later exhausted at 270°C. (a) What is the maximum efficiency that such a heat engine can have? (b) Since 270°C steam is still quite hot, a second steam engine is sometimes operated using the exhaust of the first. What is the maximum efficiency of the second engine if its exhaust has a temperature of 150°C? (c) What is the overall efficiency of the two engines? (d) Show that this is the same efficiency as a single Carnot engine operating between 450°C and 150°C. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Solution:

$$(a) \text{ } Eff_1 = 1 - \frac{T_{c,1}}{T_{h,1}} = 1 - \frac{543 \text{ K}}{723 \text{ K}} = 0.249 \text{ or } 24.9\%$$

$$(b) \text{ } Eff_2 = 1 - \frac{423 \text{ K}}{543 \text{ K}} = 0.221 \text{ or } 22.1\%$$

$$(c) \text{Eff}_1 = 1 - \frac{T_{c,1}}{T_{h,1}} \Rightarrow T_{c,1} = T_{h,1}(1 - \text{eff}_1)$$

$$\text{similarly, } T_{c,2} = T_{h,2}(1 - \text{Eff}_2)$$

using $T_{h,2} = T_{c,1}$ in above equation gives

$$T_{c,2} = T_{h,1}(1 - \text{Eff}_1)(1 - \text{Eff}_2) \equiv T_{h,1}(1 - \text{Eff}_{\text{overall}})$$

$$\therefore (1 - \text{Eff}_{\text{overall}}) = (1 - \text{Eff}_1)(1 - \text{Eff}_2)$$

$$\text{Eff}_{\text{overall}} = 1 - (1 - 0.249)(1 - 0.221) = 41.5\%$$

$$(d) \text{Eff}_{\text{overall}} = 1 - \frac{423 \text{ K}}{723 \text{ K}} = 0.415 \text{ or } 41.5\%$$

Exercise:

Problem:

A coal-fired electrical power station has an efficiency of 38%. The temperature of the steam leaving the boiler is 550°C. What percentage of the maximum efficiency does this station obtain? (Assume the temperature of the environment is 20°C.)

Exercise:

Problem:

Would you be willing to financially back an inventor who is marketing a device that she claims has 25 kJ of heat transfer at 600 K, has heat transfer to the environment at 300 K, and does 12 kJ of work? Explain your answer.

Solution:

The heat transfer to the cold reservoir is

$$Q_c = Q_h - W = 25 \text{ kJ} - 12 \text{ kJ} = 13 \text{ kJ}, \text{ so the efficiency is}$$

$$\text{Eff} = 1 - \frac{Q_c}{Q_h} = 1 - \frac{13 \text{ kJ}}{25 \text{ kJ}} = 0.48. \text{ The Carnot efficiency is}$$

$$\text{Eff}_C = 1 - \frac{T_c}{T_h} = 1 - \frac{300 \text{ K}}{600 \text{ K}} = 0.50. \text{ The actual efficiency is } 96\%$$

of the Carnot efficiency, which is much higher than the best-ever achieved of about 70%, so her scheme is likely to be fraudulent.

Exercise:

Problem: Unreasonable Results

(a) Suppose you want to design a steam engine that has heat transfer to the environment at 270°C and has a Carnot efficiency of 0.800. What temperature of hot steam must you use? (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

Exercise:

Problem: Unreasonable Results

Calculate the cold reservoir temperature of a steam engine that uses hot steam at 450°C and has a Carnot efficiency of 0.700. (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

Solution:

(a) -56.3°C

(b) The temperature is too cold for the output of a steam engine (the local environment). It is below the freezing point of water.

(c) The assumed efficiency is too high.

Glossary

Carnot cycle

a cyclical process that uses only reversible processes, the adiabatic and isothermal processes

Carnot engine

a heat engine that uses a Carnot cycle

Carnot efficiency

the maximum theoretical efficiency for a heat engine

Applications of Thermodynamics: Heat Pumps and Refrigerators

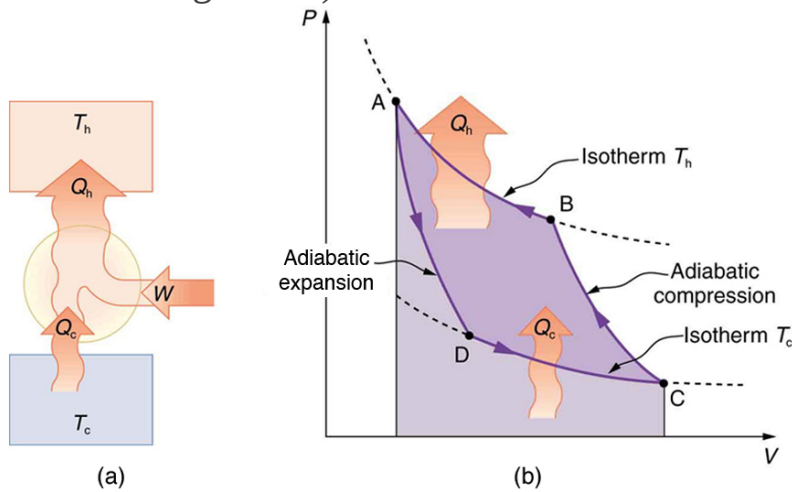
- Describe the use of heat engines in heat pumps and refrigerators.
- Demonstrate how a heat pump works to warm an interior space.
- Explain the differences between heat pumps and refrigerators.
- Calculate a heat pump's coefficient of performance.



Almost every home contains a refrigerator. Most people don't realize they are also sharing their homes with a heat pump.
(credit: Id1337x, Wikimedia Commons)

Heat pumps, air conditioners, and refrigerators utilize heat transfer from cold to hot. They are heat engines run backward. We say backward, rather than reverse, because except for Carnot engines, all heat engines, though they can be run backward, cannot truly be reversed. Heat transfer occurs from a cold reservoir Q_c and into a hot one. This requires work input W , which is also converted to heat transfer. Thus the heat transfer to the hot reservoir is $Q_h = Q_c + W$. (Note that Q_h , Q_c , and W are positive, with their directions indicated on schematics rather than by sign.) A heat pump's mission is for heat transfer Q_h to occur into a warm environment, such as a home in the winter. The mission of air conditioners and refrigerators is for

heat transfer Q_c to occur from a cool environment, such as chilling a room or keeping food at lower temperatures than the environment. (Actually, a heat pump can be used both to heat and cool a space. It is essentially an air conditioner and a heating unit all in one. In this section we will concentrate on its heating mode.)

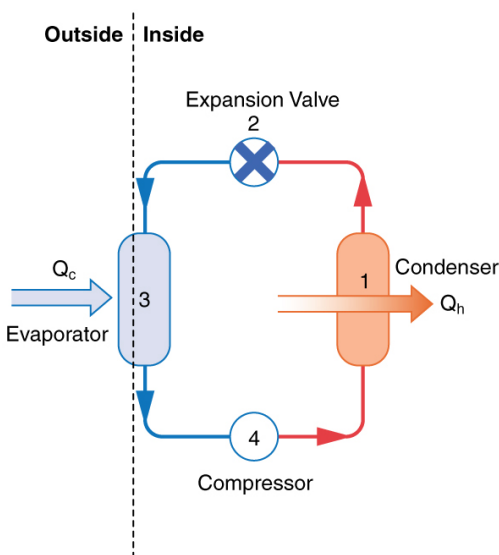


Heat pumps, air conditioners, and refrigerators are heat engines operated backward. The one shown here is based on a Carnot (reversible) engine. (a) Schematic diagram showing heat transfer from a cold reservoir to a warm reservoir with a heat pump. The directions of W , Q_h , and Q_c are opposite what they would be in a heat engine. (b) PV diagram for a Carnot cycle similar to that in [\[link\]](#) but reversed, following path ADCBA. The area inside the loop is negative, meaning there is a net work input. There is heat transfer Q_c into the system from a cold reservoir along path DC, and heat transfer Q_h out of the system into a hot reservoir along path BA.

Heat Pumps

The great advantage of using a heat pump to keep your home warm, rather than just burning fuel, is that a heat pump supplies $Q_h = Q_c + W$. Heat transfer is from the outside air, even at a temperature below freezing, to the indoor space. You only pay for W , and you get an additional heat transfer of Q_c from the outside at no cost; in many cases, at least twice as much energy is transferred to the heated space as is used to run the heat pump. When you burn fuel to keep warm, you pay for all of it. The disadvantage is that the work input (required by the second law of thermodynamics) is sometimes more expensive than simply burning fuel, especially if the work is done by electrical energy.

The basic components of a heat pump in its heating mode are shown in [\[link\]](#). A working fluid such as a non-CFC refrigerant is used. In the outdoor coils (the evaporator), heat transfer Q_c occurs to the working fluid from the cold outdoor air, turning it into a gas.



A simple heat pump has four basic components:

- (1) condenser,
- (2) expansion valve,
- (3) evaporator, and
- (4) compressor. In the

heating mode, heat transfer Q_c occurs to the working fluid in the evaporator (3) from the colder outdoor air, turning it into a gas. The electrically driven compressor (4) increases the temperature and pressure of the gas and forces it into the condenser coils (1) inside the heated space. Because the temperature of the gas is higher than the temperature in the room, heat transfer from the gas to the room occurs as the gas condenses to a liquid. The working fluid is then cooled as it flows back through an expansion valve (2) to the outdoor evaporator coils.

The electrically driven compressor (work input W) raises the temperature and pressure of the gas and forces it into the condenser coils that are inside the heated space. Because the temperature of the gas is higher than the temperature inside the room, heat transfer to the room occurs and the gas condenses to a liquid. The liquid then flows back through a pressure-reducing valve to the outdoor evaporator coils, being cooled through expansion. (In a cooling cycle, the evaporator and condenser coils exchange roles and the flow direction of the fluid is reversed.)

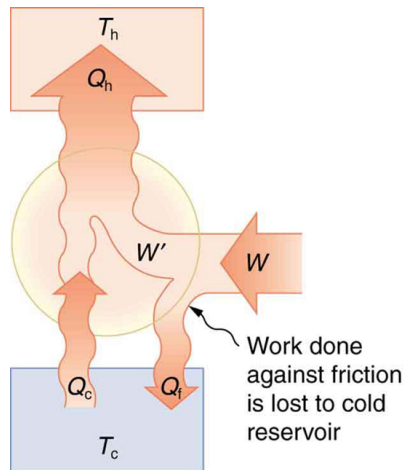
The quality of a heat pump is judged by how much heat transfer Q_h occurs into the warm space compared with how much work input W is required. In the spirit of taking the ratio of what you get to what you spend, we define a **heat pump's coefficient of performance** (COP_{hp}) to be

Equation:

$$COP_{hp} = \frac{Q_h}{W}.$$

Since the efficiency of a heat engine is $Eff = W/Q_h$, we see that $COP_{hp} = 1/Eff$, an important and interesting fact. First, since the efficiency of any heat engine is less than 1, it means that COP_{hp} is always greater than 1—that is, a heat pump always has more heat transfer Q_h than work put into it. Second, it means that heat pumps work best when temperature differences are small. The efficiency of a perfect, or Carnot, engine is $Eff_C = 1 - (T_c/T_h)$; thus, the smaller the temperature difference, the smaller the efficiency and the greater the COP_{hp} (because $COP_{hp} = 1/Eff$). In other words, heat pumps do not work as well in very cold climates as they do in more moderate climates.

Friction and other irreversible processes reduce heat engine efficiency, but they do *not* benefit the operation of a heat pump—instead, they reduce the work input by converting part of it to heat transfer back into the cold reservoir before it gets into the heat pump.



When a real heat engine is run backward, some of the intended work input (W) goes into heat transfer before it gets into the heat engine, thereby reducing its coefficient of performance COP_{hp} . In this figure, W' represents the portion of W that goes into the heat pump, while the remainder of W is lost in the form of frictional heat (Q_f) to the cold reservoir. If all of W had gone into the heat pump, then Q_h would have

been greater. The best heat pump uses adiabatic and isothermal processes, since, in theory, there would be no dissipative processes to reduce the heat transfer to the hot reservoir.

Example:

The Best COP_{hp} of a Heat Pump for Home Use

A heat pump used to warm a home must employ a cycle that produces a working fluid at temperatures greater than typical indoor temperature so that heat transfer to the inside can take place. Similarly, it must produce a working fluid at temperatures that are colder than the outdoor temperature so that heat transfer occurs from outside. Its hot and cold reservoir temperatures therefore cannot be too close, placing a limit on its COP_{hp} . (See [\[link\]](#).) What is the best coefficient of performance possible for such a heat pump, if it has a hot reservoir temperature of 45.0°C and a cold reservoir temperature of -15.0°C ?

Strategy

A Carnot engine reversed will give the best possible performance as a heat pump. As noted above, $COP_{hp} = 1/Eff$, so that we need to first calculate the Carnot efficiency to solve this problem.

Solution

Carnot efficiency in terms of absolute temperature is given by:

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

The temperatures in kelvins are $T_h = 318 \text{ K}$ and $T_c = 258 \text{ K}$, so that

Equation:

$$Eff_C = 1 - \frac{258 \text{ K}}{318 \text{ K}} = 0.1887.$$

Thus, from the discussion above,

Equation:

$$COP_{hp} = \frac{1}{Eff} = \frac{1}{0.1887} = 5.30,$$

or

Equation:

$$COP_{hp} = \frac{Q_h}{W} = 5.30,$$

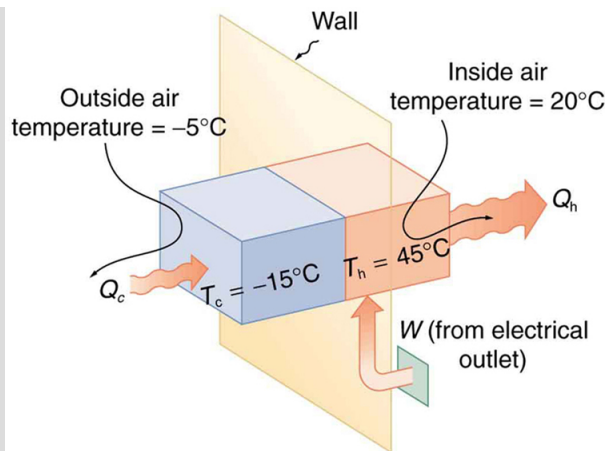
so that

Equation:

$$Q_h = 5.30 \text{ W}.$$

Discussion

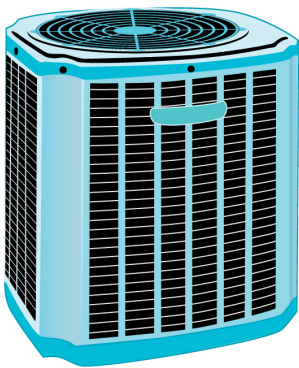
This result means that the heat transfer by the heat pump is 5.30 times as much as the work put into it. It would cost 5.30 times as much for the same heat transfer by an electric room heater as it does for that produced by this heat pump. This is not a violation of conservation of energy. Cold ambient air provides 4.3 J per 1 J of work from the electrical outlet.



Heat transfer from the outside to the inside, along with work done to run the pump, takes place in the heat pump of the example above. Note that the cold temperature produced by the heat pump is lower than the outside temperature, so that heat transfer into the working fluid occurs. The pump's compressor produces a temperature greater than the indoor temperature in order for heat transfer into the house to occur.

Real heat pumps do not perform quite as well as the ideal one in the previous example; their values of COP_{hp} range from about 2 to 4. This range means that the heat transfer Q_h from the heat pumps is 2 to 4 times as great as the work W put into them. Their economical feasibility is still limited, however, since W is usually supplied by electrical energy that costs more per joule than heat transfer by burning fuels like natural gas. Furthermore, the initial cost of a heat pump is greater than that of many

furnaces, so that a heat pump must last longer for its cost to be recovered. Heat pumps are most likely to be economically superior where winter temperatures are mild, electricity is relatively cheap, and other fuels are relatively expensive. Also, since they can cool as well as heat a space, they have advantages where cooling in summer months is also desired. Thus some of the best locations for heat pumps are in warm summer climates with cool winters. [\[link\]](#) shows a heat pump, called a “*reverse cycle*” or “*split-system cooler*” in some countries.



In hot weather, heat transfer occurs from air inside the room to air outside, cooling the room. In cool weather, heat transfer occurs from air outside to air inside, warming the room. This switching is

achieved by
reversing the
direction of
flow of the
working
fluid.

Air Conditioners and Refrigerators

Air conditioners and refrigerators are designed to cool something down in a warm environment. As with heat pumps, work input is required for heat transfer from cold to hot, and this is expensive. The quality of air conditioners and refrigerators is judged by how much heat transfer Q_c occurs from a cold environment compared with how much work input W is required. What is considered the benefit in a heat pump is considered waste heat in a refrigerator. We thus define the **coefficient of performance** (COP_{ref}) of an air conditioner or refrigerator to be

Equation:

$$COP_{\text{ref}} = \frac{Q_c}{W}.$$

Noting again that $Q_h = Q_c + W$, we can see that an air conditioner will have a lower coefficient of performance than a heat pump, because $COP_{\text{hp}} = Q_h/W$ and Q_h is greater than Q_c . In this module's Problems and Exercises, you will show that

Equation:

$$COP_{\text{ref}} = COP_{\text{hp}} - 1$$

for a heat engine used as either an air conditioner or a heat pump operating between the same two temperatures. Real air conditioners and refrigerators typically do remarkably well, having values of COP_{ref} ranging from 2 to 6.

These numbers are better than the COP_{hp} values for the heat pumps mentioned above, because the temperature differences are smaller, but they are less than those for Carnot engines operating between the same two temperatures.

A type of COP rating system called the “energy efficiency rating” (EER) has been developed. This rating is an example where non-SI units are still used and relevant to consumers. To make it easier for the consumer, Australia, Canada, New Zealand, and the U.S. use an Energy Star Rating out of 5 stars—the more stars, the more energy efficient the appliance. EER s are expressed in mixed units of British thermal units (Btu) per hour of heating or cooling divided by the power input in watts. Room air conditioners are readily available with EER s ranging from 6 to 12. Although not the same as the COP s just described, these EER s are good for comparison purposes—the greater the EER , the cheaper an air conditioner is to operate (but the higher its purchase price is likely to be).

The EER of an air conditioner or refrigerator can be expressed as

Equation:

$$EER = \frac{Q_c/t_1}{W/t_2},$$

where Q_c is the amount of heat transfer from a cold environment in British thermal units, t_1 is time in hours, W is the work input in joules, and t_2 is time in seconds.

Note:

Problem-Solving Strategies for Thermodynamics

1. *Examine the situation to determine whether heat, work, or internal energy are involved.* Look for any system where the primary methods of transferring energy are heat and work. Heat engines, heat pumps, refrigerators, and air conditioners are examples of such systems.

2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Maximum efficiency means a Carnot engine is involved. Efficiency is not the same as the coefficient of performance.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns). Be sure to distinguish heat transfer into a system from heat transfer out of the system, as well as work input from work output. In many situations, it is useful to determine the type of process, such as isothermal or adiabatic.*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. *Substitute the known quantities along with their units into the appropriate equation and obtain numerical solutions complete with units.*
7. *Check the answer to see if it is reasonable: Does it make sense? For example, efficiency is always less than 1, whereas coefficients of performance are greater than 1.*

Section Summary

- An artifact of the second law of thermodynamics is the ability to heat an interior space using a heat pump. Heat pumps compress cold ambient air and, in so doing, heat it to room temperature without violation of conservation principles.
- To calculate the heat pump's coefficient of performance, use the equation $COP_{hp} = \frac{Q_h}{W}$.
- A refrigerator is a heat pump; it takes warm ambient air and expands it to chill it.

Conceptual Questions

Exercise:

Problem:

Explain why heat pumps do not work as well in very cold climates as they do in milder ones. Is the same true of refrigerators?

Exercise:**Problem:**

In some Northern European nations, homes are being built without heating systems of any type. They are very well insulated and are kept warm by the body heat of the residents. However, when the residents are not at home, it is still warm in these houses. What is a possible explanation?

Exercise:**Problem:**

Why do refrigerators, air conditioners, and heat pumps operate most cost-effectively for cycles with a small difference between T_h and T_c ? (Note that the temperatures of the cycle employed are crucial to its COP.)

Exercise:**Problem:**

Grocery store managers contend that there is *less* total energy consumption in the summer if the store is kept at a *low* temperature. Make arguments to support or refute this claim, taking into account that there are numerous refrigerators and freezers in the store.

Exercise:**Problem:**

Can you cool a kitchen by leaving the refrigerator door open?

Problem Exercises

Exercise:**Problem:**

What is the coefficient of performance of an ideal heat pump that has heat transfer from a cold temperature of -25.0°C to a hot temperature of 40.0°C ?

Solution:

4.82

Exercise:**Problem:**

Suppose you have an ideal refrigerator that cools an environment at -20.0°C and has heat transfer to another environment at 50.0°C . What is its coefficient of performance?

Exercise:**Problem:**

What is the best coefficient of performance possible for a hypothetical refrigerator that could make liquid nitrogen at -200°C and has heat transfer to the environment at 35.0°C ?

Solution:

0.311

Exercise:**Problem:**

In a very mild winter climate, a heat pump has heat transfer from an environment at 5.00°C to one at 35.0°C . What is the best possible coefficient of performance for these temperatures? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Exercise:**Problem:**

(a) What is the best coefficient of performance for a heat pump that has a hot reservoir temperature of 50.0°C and a cold reservoir temperature of -20.0°C ? (b) How much heat transfer occurs into the warm environment if $3.60 \times 10^7 \text{ J}$ of work ($10.0 \text{ kW} \cdot \text{h}$) is put into it? (c) If the cost of this work input is $10.0 \text{ cents/kW} \cdot \text{h}$, how does its cost compare with the direct heat transfer achieved by burning natural gas at a cost of $85.0 \text{ cents per therm}$. (A therm is a common unit of energy for natural gas and equals $1.055 \times 10^8 \text{ J}$.)

Solution:

(a) 4.61

(b) $1.66 \times 10^8 \text{ J}$ or $3.97 \times 10^4 \text{ kcal}$

(c) To transfer $1.66 \times 10^8 \text{ J}$, heat pump costs \$1.00, natural gas costs \$1.34.

Exercise:**Problem:**

(a) What is the best coefficient of performance for a refrigerator that cools an environment at -30.0°C and has heat transfer to another environment at 45.0°C ? (b) How much work in joules must be done for a heat transfer of 4186 kJ from the cold environment? (c) What is the cost of doing this if the work costs $10.0 \text{ cents per } 3.60 \times 10^6 \text{ J}$ (a kilowatt-hour)? (d) How many kJ of heat transfer occurs into the warm environment? (e) Discuss what type of refrigerator might operate between these temperatures.

Exercise:

Problem:

Suppose you want to operate an ideal refrigerator with a cold temperature of -10.0°C , and you would like it to have a coefficient of performance of 7.00. What is the hot reservoir temperature for such a refrigerator?

Solution:

27.6°C

Exercise:**Problem:**

An ideal heat pump is being considered for use in heating an environment with a temperature of 22.0°C . What is the cold reservoir temperature if the pump is to have a coefficient of performance of 12.0?

Exercise:**Problem:**

A 4-ton air conditioner removes 5.06×10^7 J (48,000 British thermal units) from a cold environment in 1.00 h. (a) What energy input in joules is necessary to do this if the air conditioner has an energy efficiency rating (*EER*) of 12.0? (b) What is the cost of doing this if the work costs 10.0 cents per 3.60×10^6 J (one kilowatt-hour)? (c) Discuss whether this cost seems realistic. Note that the energy efficiency rating (*EER*) of an air conditioner or refrigerator is defined to be the number of British thermal units of heat transfer from a cold environment per hour divided by the watts of power input.

Solution:

(a) 1.44×10^7 J

(b) 40 cents

(c) This cost seems quite realistic; it says that running an air conditioner all day would cost \$9.59 (if it ran continuously).

Exercise:

Problem:

Show that the coefficients of performance of refrigerators and heat pumps are related by $COP_{\text{ref}} = COP_{\text{hp}} - 1$.

Start with the definitions of the COP s and the conservation of energy relationship between Q_h , Q_c , and W .

Glossary

heat pump

a machine that generates heat transfer from cold to hot

coefficient of performance

for a heat pump, it is the ratio of heat transfer at the output (the hot reservoir) to the work supplied; for a refrigerator or air conditioner, it is the ratio of heat transfer from the cold reservoir to the work supplied

Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy

- Define entropy and calculate the increase of entropy in a system with reversible and irreversible processes.
- Explain the expected fate of the universe in entropic terms.
- Calculate the increasing disorder of a system.



The ice in this drink is slowly melting. Eventually the liquid will reach thermal equilibrium, as predicted by the second law of thermodynamics.

(credit: Jon Sullivan, PDPhoto.org)

There is yet another way of expressing the second law of thermodynamics. This version relates to a concept called **entropy**. By examining it, we shall

see that the directions associated with the second law—heat transfer from hot to cold, for example—are related to the tendency in nature for systems to become disordered and for less energy to be available for use as work. The entropy of a system can in fact be shown to be a measure of its disorder and of the unavailability of energy to do work.

Note:

Making Connections: Entropy, Energy, and Work

Recall that the simple definition of energy is the ability to do work. Entropy is a measure of how much energy is not available to do work. Although all forms of energy are interconvertible, and all can be used to do work, it is not always possible, even in principle, to convert the entire available energy into work. That unavailable energy is of interest in thermodynamics, because the field of thermodynamics arose from efforts to convert heat to work.

We can see how entropy is defined by recalling our discussion of the Carnot engine. We noted that for a Carnot cycle, and hence for any reversible processes, $Q_c/Q_h = T_c/T_h$. Rearranging terms yields

Equation:

$$\frac{Q_c}{T_c} = \frac{Q_h}{T_h}$$

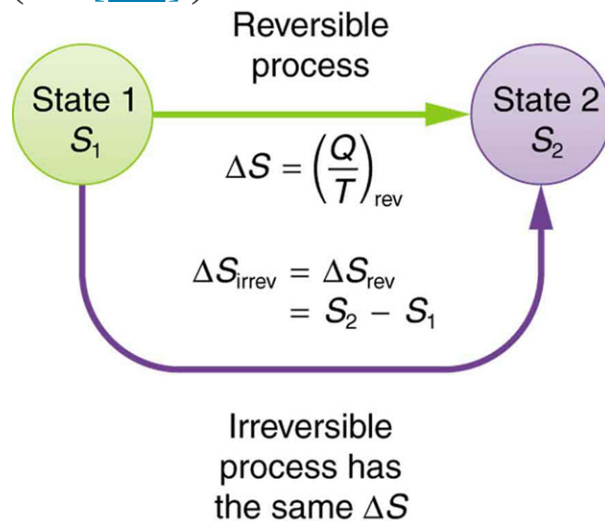
for any reversible process. Q_c and Q_h are absolute values of the heat transfer at temperatures T_c and T_h , respectively. This ratio of Q/T is defined to be the **change in entropy** ΔS for a reversible process,

Equation:

$$\Delta S = \left(\frac{Q}{T} \right)_{\text{rev}},$$

where Q is the heat transfer, which is positive for heat transfer into and negative for heat transfer out of, and T is the absolute temperature at which the reversible process takes place. The SI unit for entropy is joules per kelvin (J/K). If temperature changes during the process, then it is usually a good approximation (for small changes in temperature) to take T to be the average temperature, avoiding the need to use integral calculus to find ΔS .

The definition of ΔS is strictly valid only for reversible processes, such as used in a Carnot engine. However, we can find ΔS precisely even for real, irreversible processes. The reason is that the entropy S of a system, like internal energy U , depends only on the state of the system and not how it reached that condition. Entropy is a property of state. Thus the change in entropy ΔS of a system between state 1 and state 2 is the same no matter how the change occurs. We just need to find or imagine a reversible process that takes us from state 1 to state 2 and calculate ΔS for that process. That will be the change in entropy for any process going from state 1 to state 2. (See [\[link\]](#).)



When a system goes from state 1 to state 2, its entropy changes by the same amount ΔS , whether a hypothetical reversible path is followed or a real irreversible path is taken.

Now let us take a look at the change in entropy of a Carnot engine and its heat reservoirs for one full cycle. The hot reservoir has a loss of entropy $\Delta S_h = -Q_h/T_h$, because heat transfer occurs out of it (remember that when heat transfers out, then Q has a negative sign). The cold reservoir has a gain of entropy $\Delta S_c = Q_c/T_c$, because heat transfer occurs into it. (We assume the reservoirs are sufficiently large that their temperatures are constant.) So the total change in entropy is

Equation:

$$\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c.$$

Thus, since we know that $Q_h/T_h = Q_c/T_c$ for a Carnot engine,

Equation:

$$\Delta S_{\text{tot}} = -\frac{Q_h}{T_h} + \frac{Q_c}{T_c} = 0.$$

This result, which has general validity, means that *the total change in entropy for a system in any reversible process is zero.*

The entropy of various parts of the system may change, but the total change is zero. Furthermore, the system does not affect the entropy of its surroundings, since heat transfer between them does not occur. Thus the reversible process changes neither the total entropy of the system nor the entropy of its surroundings. Sometimes this is stated as follows: *Reversible processes do not affect the total entropy of the universe.* Real processes are not reversible, though, and they do change total entropy. We can, however, use hypothetical reversible processes to determine the value of entropy in real, irreversible processes. The following example illustrates this point.

Example:

Entropy Increases in an Irreversible (Real) Process

Spontaneous heat transfer from hot to cold is an irreversible process. Calculate the total change in entropy if 4000 J of heat transfer occurs from

a hot reservoir at $T_h = 600 \text{ K}$ (327° C) to a cold reservoir at $T_c = 250 \text{ K}$ (-23° C), assuming there is no temperature change in either reservoir. (See [\[link\]](#).)

Strategy

How can we calculate the change in entropy for an irreversible process when $\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c$ is valid only for reversible processes? Remember that the total change in entropy of the hot and cold reservoirs will be the same whether a reversible or irreversible process is involved in heat transfer from hot to cold. So we can calculate the change in entropy of the hot reservoir for a hypothetical reversible process in which 4000 J of heat transfer occurs from it; then we do the same for a hypothetical reversible process in which 4000 J of heat transfer occurs to the cold reservoir. This produces the same changes in the hot and cold reservoirs that would occur if the heat transfer were allowed to occur irreversibly between them, and so it also produces the same changes in entropy.

Solution

We now calculate the two changes in entropy using $\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c$. First, for the heat transfer from the hot reservoir,

Equation:

$$\Delta S_h = \frac{-Q_h}{T_h} = \frac{-4000 \text{ J}}{600 \text{ K}} = -6.67 \text{ J/K}.$$

And for the cold reservoir,

Equation:

$$\Delta S_c = \frac{Q_c}{T_c} = \frac{4000 \text{ J}}{250 \text{ K}} = 16.0 \text{ J/K}.$$

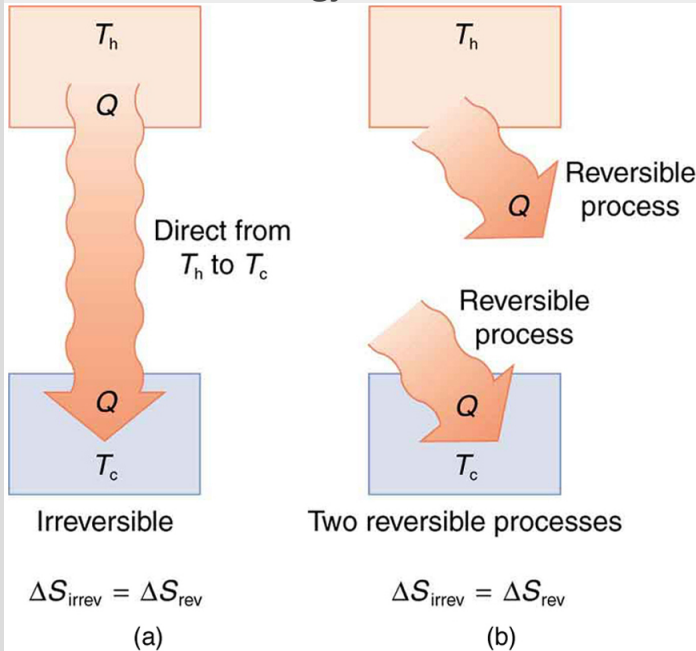
Thus the total is

Equation:

$$\begin{aligned}\Delta S_{\text{tot}} &= \Delta S_h + \Delta S_c \\ &= (-6.67 + 16.0) \text{ J/K} \\ &= 9.33 \text{ J/K}.\end{aligned}$$

Discussion

There is an *increase* in entropy for the system of two heat reservoirs undergoing this irreversible heat transfer. We will see that this means there is a loss of ability to do work with this transferred energy. Entropy has increased, and energy has become unavailable to do work.



(a) Heat transfer from a hot object to a cold one is an irreversible process that produces an overall increase in entropy. (b) The same final state and, thus, the same change in entropy is achieved for the objects if reversible heat transfer processes occur between the two objects whose temperatures are the same as the temperatures of the corresponding objects in the irreversible process.

It is reasonable that entropy increases for heat transfer from hot to cold. Since the change in entropy is Q/T , there is a larger change at lower

temperatures. The decrease in entropy of the hot object is therefore less than the increase in entropy of the cold object, producing an overall increase, just as in the previous example. This result is very general:

There is an increase in entropy for any system undergoing an irreversible process.

With respect to entropy, there are only two possibilities: entropy is constant for a reversible process, and it increases for an irreversible process. There is a fourth version of **the second law of thermodynamics stated in terms of entropy**:

The total entropy of a system either increases or remains constant in any process; it never decreases.

For example, heat transfer cannot occur spontaneously from cold to hot, because entropy would decrease.

Entropy is very different from energy. Entropy is *not* conserved but increases in all real processes. Reversible processes (such as in Carnot engines) are the processes in which the most heat transfer to work takes place and are also the ones that keep entropy constant. Thus we are led to make a connection between entropy and the availability of energy to do work.

Entropy and the Unavailability of Energy to Do Work

What does a change in entropy mean, and why should we be interested in it? One reason is that entropy is directly related to the fact that not all heat transfer can be converted into work. The next example gives some indication of how an increase in entropy results in less heat transfer into work.

Example:

Less Work is Produced by a Given Heat Transfer When Entropy Change is Greater

(a) Calculate the work output of a Carnot engine operating between temperatures of 600 K and 100 K for 4000 J of heat transfer to the engine.
(b) Now suppose that the 4000 J of heat transfer occurs first from the 600 K reservoir to a 250 K reservoir (without doing any work, and this produces the increase in entropy calculated above) before transferring into a Carnot engine operating between 250 K and 100 K. What work output is produced? (See [\[link\]](#).)

Strategy

In both parts, we must first calculate the Carnot efficiency and then the work output.

Solution (a)

The Carnot efficiency is given by

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

Substituting the given temperatures yields

Equation:

$$Eff_C = 1 - \frac{100 \text{ K}}{600 \text{ K}} = 0.833.$$

Now the work output can be calculated using the definition of efficiency for any heat engine as given by

Equation:

$$Eff = \frac{W}{Q_h}.$$

Solving for W and substituting known terms gives

Equation:

$$\begin{aligned} W &= Eff_C Q_h \\ &= (0.833)(4000 \text{ J}) = 3333 \text{ J}. \end{aligned}$$

Solution (b)

Similarly,

Equation:

$$Eff'_C = 1 - \frac{T_c}{T'_c} = 1 - \frac{100 \text{ K}}{250 \text{ K}} = 0.600,$$

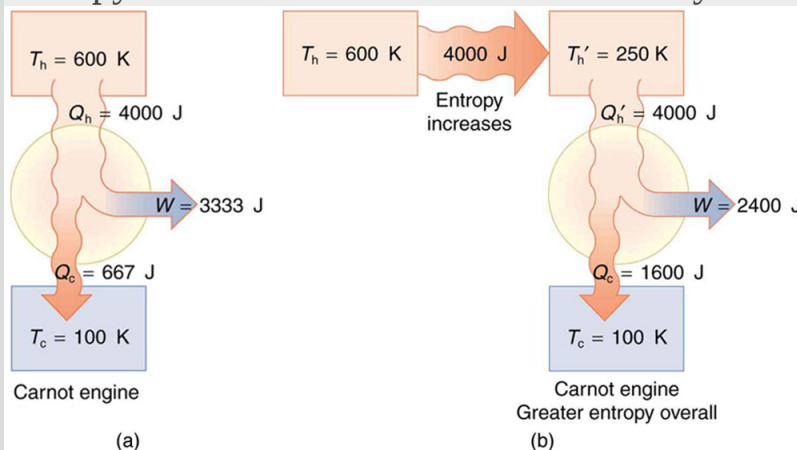
so that

Equation:

$$\begin{aligned} W &= Eff'_C Q_h \\ &= (0.600)(4000 \text{ J}) = 2400 \text{ J}. \end{aligned}$$

Discussion

There is 933 J less work from the same heat transfer in the second process. This result is important. The same heat transfer into two perfect engines produces different work outputs, because the entropy change differs in the two cases. In the second case, entropy is greater and less work is produced. Entropy is associated with the *unavailability* of energy to do work.



(a) A Carnot engine working at between 600 K and 100 K has 4000 J of heat transfer and performs 3333 J of work. (b) The 4000 J of heat transfer occurs first irreversibly to a 250 K reservoir and then goes into a Carnot engine. The increase in entropy caused by the heat transfer to a colder reservoir results in a smaller work output of 2400 J. There is a permanent loss of 933 J of energy for the purpose of doing work.

When entropy increases, a certain amount of energy becomes *permanently* unavailable to do work. The energy is not lost, but its character is changed, so that some of it can never be converted to doing work—that is, to an organized force acting through a distance. For instance, in the previous example, 933 J less work was done after an increase in entropy of 9.33 J/K occurred in the 4000 J heat transfer from the 600 K reservoir to the 250 K reservoir. It can be shown that the amount of energy that becomes unavailable for work is

Equation:

$$W_{\text{unavail}} = \Delta S \cdot T_0,$$

where T_0 is the lowest temperature utilized. In the previous example,

Equation:

$$W_{\text{unavail}} = (9.33 \text{ J/K})(100 \text{ K}) = 933 \text{ J}$$

as found.

Heat Death of the Universe: An Overdose of Entropy

In the early, energetic universe, all matter and energy were easily interchangeable and identical in nature. Gravity played a vital role in the young universe. Although it may have *seemed* disorderly, and therefore, superficially entropic, in fact, there was enormous potential energy available to do work—all the future energy in the universe.

As the universe matured, temperature differences arose, which created more opportunity for work. Stars are hotter than planets, for example, which are warmer than icy asteroids, which are warmer still than the vacuum of the space between them.

Most of these are cooling down from their usually violent births, at which time they were provided with energy of their own—nuclear energy in the case of stars, volcanic energy on Earth and other planets, and so on. Without additional energy input, however, their days are numbered.

As entropy increases, less and less energy in the universe is available to do work. On Earth, we still have great stores of energy such as fossil and nuclear fuels; large-scale temperature differences, which can provide wind energy; geothermal energies due to differences in temperature in Earth's layers; and tidal energies owing to our abundance of liquid water. As these are used, a certain fraction of the energy they contain can never be converted into doing work. Eventually, all fuels will be exhausted, all temperatures will equalize, and it will be impossible for heat engines to function, or for work to be done.

Entropy increases in a closed system, such as the universe. But in parts of the universe, for instance, in the Solar system, it is not a locally closed system. Energy flows from the Sun to the planets, replenishing Earth's stores of energy. The Sun will continue to supply us with energy for about another five billion years. We will enjoy direct solar energy, as well as side effects of solar energy, such as wind power and biomass energy from photosynthetic plants. The energy from the Sun will keep our water at the liquid state, and the Moon's gravitational pull will continue to provide tidal energy. But Earth's geothermal energy will slowly run down and won't be replenished.

But in terms of the universe, and the very long-term, very large-scale picture, the entropy of the universe is increasing, and so the availability of energy to do work is constantly decreasing. Eventually, when all stars have died, all forms of potential energy have been utilized, and all temperatures have equalized (depending on the mass of the universe, either at a very high temperature following a universal contraction, or a very low one, just before all activity ceases) there will be no possibility of doing work.

Either way, the universe is destined for thermodynamic equilibrium—maximum entropy. This is often called the *heat death of the universe*, and will mean the end of all activity. However, whether the universe contracts and heats up, or continues to expand and cools down, the end is not near.

Calculations of black holes suggest that entropy can easily continue for at least 10^{100} years.

Order to Disorder

Entropy is related not only to the unavailability of energy to do work—it is also a measure of disorder. This notion was initially postulated by Ludwig Boltzmann in the 1800s. For example, melting a block of ice means taking a highly structured and orderly system of water molecules and converting it into a disorderly liquid in which molecules have no fixed positions. (See [\[link\]](#).) There is a large increase in entropy in the process, as seen in the following example.

Example:

Entropy Associated with Disorder

Find the increase in entropy of 1.00 kg of ice originally at 0°C that is melted to form water at 0°C .

Strategy

As before, the change in entropy can be calculated from the definition of ΔS once we find the energy Q needed to melt the ice.

Solution

The change in entropy is defined as:

Equation:

$$\Delta S = \frac{Q}{T}.$$

Here Q is the heat transfer necessary to melt 1.00 kg of ice and is given by

Equation:

$$Q = mL_f,$$

where m is the mass and L_f is the latent heat of fusion. $L_f = 334\text{ kJ/kg}$ for water, so that

Equation:

$$Q = (1.00 \text{ kg})(334 \text{ kJ/kg}) = 3.34 \times 10^5 \text{ J}.$$

Now the change in entropy is positive, since heat transfer occurs into the ice to cause the phase change; thus,

Equation:

$$\Delta S = \frac{Q}{T} = \frac{3.34 \times 10^5 \text{ J}}{T}.$$

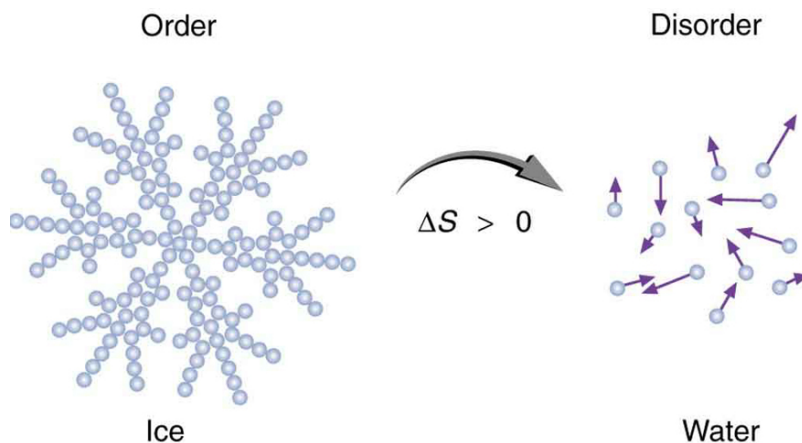
T is the melting temperature of ice. That is, $T = 0^\circ\text{C} = 273 \text{ K}$. So the change in entropy is

Equation:

$$\begin{aligned} \Delta S &= \frac{3.34 \times 10^5 \text{ J}}{273 \text{ K}} \\ &= 1.22 \times 10^3 \text{ J/K}. \end{aligned}$$

Discussion

This is a significant increase in entropy accompanying an increase in disorder.



When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without

fixed locations or orientations. Its entropy increases because heat transfer occurs into it. Entropy is a measure of disorder.

In another easily imagined example, suppose we mix equal masses of water originally at two different temperatures, say 20.0°C and 40.0°C . The result is water at an intermediate temperature of 30.0°C . Three outcomes have resulted: entropy has increased, some energy has become unavailable to do work, and the system has become less orderly. Let us think about each of these results.

First, entropy has increased for the same reason that it did in the example above. Mixing the two bodies of water has the same effect as heat transfer from the hot one and the same heat transfer into the cold one. The mixing decreases the entropy of the hot water but increases the entropy of the cold water by a greater amount, producing an overall increase in entropy.

Second, once the two masses of water are mixed, there is only one temperature—you cannot run a heat engine with them. The energy that could have been used to run a heat engine is now unavailable to do work.

Third, the mixture is less orderly, or to use another term, less structured. Rather than having two masses at different temperatures and with different distributions of molecular speeds, we now have a single mass with a uniform temperature.

These three results—entropy, unavailability of energy, and disorder—are not only related but are in fact essentially equivalent.

Life, Evolution, and the Second Law of Thermodynamics

Some people misunderstand the second law of thermodynamics, stated in terms of entropy, to say that the process of the evolution of life violates this law. Over time, complex organisms evolved from much simpler ancestors, representing a large decrease in entropy of the Earth's biosphere. It is a fact

that living organisms have evolved to be highly structured, and much lower in entropy than the substances from which they grow. But it is *always* possible for the entropy of one part of the universe to decrease, provided the total change in entropy of the universe increases. In equation form, we can write this as

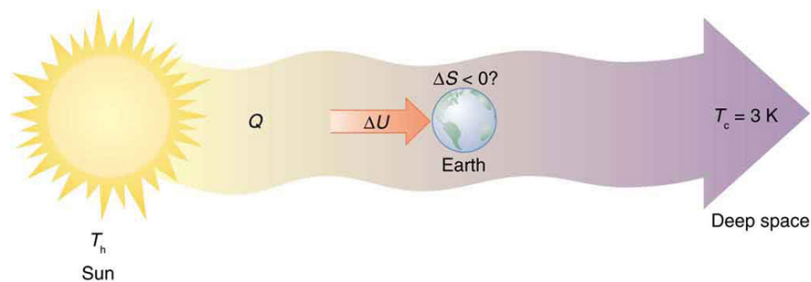
Equation:

$$\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0.$$

Thus ΔS_{syst} can be negative as long as ΔS_{envir} is positive and greater in magnitude.

How is it possible for a system to decrease its entropy? Energy transfer is necessary. If I pick up marbles that are scattered about the room and put them into a cup, my work has decreased the entropy of that system. If I gather iron ore from the ground and convert it into steel and build a bridge, my work has decreased the entropy of that system. Energy coming from the Sun can decrease the entropy of local systems on Earth—that is, ΔS_{syst} is negative. But the overall entropy of the rest of the universe increases by a greater amount—that is, ΔS_{envir} is positive and greater in magnitude. Thus, $\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0$, and the second law of thermodynamics is *not* violated.

Every time a plant stores some solar energy in the form of chemical potential energy, or an updraft of warm air lifts a soaring bird, the Earth can be viewed as a heat engine operating between a hot reservoir supplied by the Sun and a cold reservoir supplied by dark outer space—a heat engine of high complexity, causing local decreases in entropy as it uses part of the heat transfer from the Sun into deep space. There is a large total increase in entropy resulting from this massive heat transfer. A small part of this heat transfer is stored in structured systems on Earth, producing much smaller local decreases in entropy. (See [\[link\]](#).)



Earth's entropy may decrease in the process of intercepting a small part of the heat transfer from the Sun into deep space.

Entropy for the entire process increases greatly while Earth becomes more structured with living systems and stored energy in various forms.

Note:

PhET Explorations: Reversible Reactions

Watch a reaction proceed over time. How does total energy affect a reaction rate? Vary temperature, barrier height, and potential energies. Record concentrations and time in order to extract rate coefficients. Do temperature dependent studies to extract Arrhenius parameters. This simulation is best used with teacher guidance because it presents an analogy of chemical reactions.

[Reversible Reactions](#)

Section Summary

- Entropy is the loss of energy available to do work.
- Another form of the second law of thermodynamics states that the total entropy of a system either increases or remains constant; it never decreases.
- Entropy is zero in a reversible process; it increases in an irreversible process.
- The ultimate fate of the universe is likely to be thermodynamic equilibrium, where the universal temperature is constant and no energy is available to do work.
- Entropy is also associated with the tendency toward disorder in a closed system.

Conceptual Questions

Exercise:

Problem:

A woman shuts her summer cottage up in September and returns in June. No one has entered the cottage in the meantime. Explain what she is likely to find, in terms of the second law of thermodynamics.

Exercise:

Problem:

Consider a system with a certain energy content, from which we wish to extract as much work as possible. Should the system's entropy be high or low? Is this orderly or disorderly? Structured or uniform? Explain briefly.

Exercise:

Problem:

Does a gas become more orderly when it liquefies? Does its entropy change? If so, does the entropy increase or decrease? Explain your answer.

Exercise:

Problem:

Explain how water's entropy can decrease when it freezes without violating the second law of thermodynamics. Specifically, explain what happens to the entropy of its surroundings.

Exercise:

Problem:

Is a uniform-temperature gas more or less orderly than one with several different temperatures? Which is more structured? In which can heat transfer result in work done without heat transfer from another system?

Exercise:

Problem:

Give an example of a spontaneous process in which a system becomes less ordered and energy becomes less available to do work. What happens to the system's entropy in this process?

Exercise:

Problem:

What is the change in entropy in an adiabatic process? Does this imply that adiabatic processes are reversible? Can a process be precisely adiabatic for a macroscopic system?

Exercise:

Problem:

Does the entropy of a star increase or decrease as it radiates? Does the entropy of the space into which it radiates (which has a temperature of about 3 K) increase or decrease? What does this do to the entropy of the universe?

Exercise:

Problem:

Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problem Exercises**Exercise:****Problem:**

(a) On a winter day, a certain house loses 5.00×10^8 J of heat to the outside (about 500,000 Btu). What is the total change in entropy due to this heat transfer alone, assuming an average indoor temperature of 21.0°C and an average outdoor temperature of 5.00°C ? (b) This large change in entropy implies a large amount of energy has become unavailable to do work. Where do we find more energy when such energy is lost to us?

Solution:

(a) 9.78×10^4 J/K

(b) In order to gain more energy, we must generate it from things within the house, like a heat pump, human bodies, and other appliances. As you know, we use a lot of energy to keep our houses warm in the winter because of the loss of heat to the outside.

Exercise:**Problem:**

On a hot summer day, 4.00×10^6 J of heat transfer into a parked car takes place, increasing its temperature from 35.0°C to 45.0°C . What is the increase in entropy of the car due to this heat transfer alone?

Exercise:

Problem:

A hot rock ejected from a volcano's lava fountain cools from 1100°C to 40.0°C , and its entropy decreases by 950 J/K . How much heat transfer occurs from the rock?

Solution:

$$8.01 \times 10^5\text{ J}$$

Exercise:**Problem:**

When $1.60 \times 10^5\text{ J}$ of heat transfer occurs into a meat pie initially at 20.0°C , its entropy increases by 480 J/K . What is its final temperature?

Exercise:**Problem:**

The Sun radiates energy at the rate of $3.80 \times 10^{26}\text{ W}$ from its 5500°C surface into dark empty space (a negligible fraction radiates onto Earth and the other planets). The effective temperature of deep space is -270°C . (a) What is the increase in entropy in one day due to this heat transfer? (b) How much work is made unavailable?

Solution:

(a) $1.04 \times 10^{31}\text{ J/K}$

(b) $3.28 \times 10^{31}\text{ J}$

Exercise:

Problem:

(a) In reaching equilibrium, how much heat transfer occurs from 1.00 kg of water at 40.0°C when it is placed in contact with 1.00 kg of 20.0°C water in reaching equilibrium? (b) What is the change in entropy due to this heat transfer? (c) How much work is made unavailable, taking the lowest temperature to be 20.0°C ? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Entropy](#).

Exercise:**Problem:**

What is the decrease in entropy of 25.0 g of water that condenses on a bathroom mirror at a temperature of 35.0°C , assuming no change in temperature and given the latent heat of vaporization to be 2450 kJ/kg?

Solution:

199 J/K

Exercise:**Problem:**

Find the increase in entropy of 1.00 kg of liquid nitrogen that starts at its boiling temperature, boils, and warms to 20.0°C at constant pressure.

Exercise:

Problem:

A large electrical power station generates 1000 MW of electricity with an efficiency of 35.0%. (a) Calculate the heat transfer to the power station, Q_h , in one day. (b) How much heat transfer Q_c occurs to the environment in one day? (c) If the heat transfer in the cooling towers is from 35.0° C water into the local air mass, which increases in temperature from 18.0° C to 20.0° C, what is the total increase in entropy due to this heat transfer? (d) How much energy becomes unavailable to do work because of this increase in entropy, assuming an 18.0° C lowest temperature? (Part of Q_c could be utilized to operate heat engines or for simply heating the surroundings, but it rarely is.)

Solution:

(a) $2.47 \times 10^{14} \text{ J}$

(b) $1.60 \times 10^{14} \text{ J}$

(c) $2.85 \times 10^{10} \text{ J/K}$

(d) $8.29 \times 10^{12} \text{ J}$

Exercise:

Problem:

(a) How much heat transfer occurs from 20.0 kg of 90.0° C water placed in contact with 20.0 kg of 10.0° C water, producing a final temperature of 50.0° C? (b) How much work could a Carnot engine do with this heat transfer, assuming it operates between two reservoirs at constant temperatures of 90.0° C and 10.0° C? (c) What increase in entropy is produced by mixing 20.0 kg of 90.0° C water with 20.0 kg of 10.0° C water? (d) Calculate the amount of work made unavailable by this mixing using a low temperature of 10.0° C, and compare it with the work done by the Carnot engine. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Entropy](#). (e) Discuss how everyday processes make increasingly more energy unavailable to do work, as implied by this problem.

Glossary**entropy**

a measurement of a system's disorder and its inability to do work in a system

change in entropy

the ratio of heat transfer to temperature Q/T

second law of thermodynamics stated in terms of entropy

the total entropy of a system either increases or remains constant; it never decreases

Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation

- Identify probabilities in entropy.
- Analyze statistical probabilities in entropic systems.



When you toss a coin a large number of times, heads and tails tend to come up in roughly equal numbers. Why doesn't heads come up 100, 90, or even 80% of the time? (credit: Jon Sullivan, PDPhoto.org)

The various ways of formulating the second law of thermodynamics tell what happens rather than why it happens. Why should heat transfer occur only from hot to cold? Why should energy become ever less available to do work? Why should the universe become increasingly disorderly? The answer is that it is a matter of overwhelming probability. Disorder is simply vastly more likely than order.

When you watch an emerging rain storm begin to wet the ground, you will notice that the drops fall in a disorganized manner both in time and in space. Some fall close together, some far apart, but they never fall in

straight, orderly rows. It is not impossible for rain to fall in an orderly pattern, just highly unlikely, because there are many more disorderly ways than orderly ones. To illustrate this fact, we will examine some random processes, starting with coin tosses.

Coin Tosses

What are the possible outcomes of tossing 5 coins? Each coin can land either heads or tails. On the large scale, we are concerned only with the total heads and tails and not with the order in which heads and tails appear. The following possibilities exist:

Equation:

5 heads, 0 tails

4 heads, 1 tail

3 heads, 2 tails

2 heads, 3 tails

1 head, 4 tails

0 head, 5 tails

These are what we call macrostates. A **macrostate** is an overall property of a system. It does not specify the details of the system, such as the order in which heads and tails occur or which coins are heads or tails.

Using this nomenclature, a system of 5 coins has the 6 possible macrostates just listed. Some macrostates are more likely to occur than others. For instance, there is only one way to get 5 heads, but there are several ways to get 3 heads and 2 tails, making the latter macrostate more probable. [\[link\]](#) lists of all the ways in which 5 coins can be tossed, taking into account the order in which heads and tails occur. Each sequence is called a **microstate**—a detailed description of every element of a system.

	Individual microstates	Number of microstates
5 heads, 0 tails	HHHHH	1
4 heads, 1 tail	HHHHT, HHHTH, HHTHH, HTHHH, THHHH	5
3 heads, 2 tails	HTHTH, THTHH, HTHHT, THHTH, THHHT HTHTH, THTHH, HTHHT, THHTH, THHHT	10
2 heads, 3 tails	TTTHH, TTHHT, THHTT, HHTTT, TTHTH, THTHT, HTHTT, THTTH, HTTHT, HTTTH	10
1 head, 4 tails	TTTTH, TTTHT, TTHTT, THTTT, HTTTT	5
0 heads, 5 tails	TTTTT	1
		Total: 32

5-Coin Toss

The macrostate of 3 heads and 2 tails can be achieved in 10 ways and is thus 10 times more probable than the one having 5 heads. Not surprisingly, it is equally probable to have the reverse, 2 heads and 3 tails. Similarly, it is equally probable to get 5 tails as it is to get 5 heads. Note that all of these conclusions are based on the crucial assumption that each microstate is equally probable. With coin tosses, this requires that the coins not be

asymmetric in a way that favors one side over the other, as with loaded dice. With any system, the assumption that all microstates are equally probable must be valid, or the analysis will be erroneous.

The two most orderly possibilities are 5 heads or 5 tails. (They are more structured than the others.) They are also the least likely, only 2 out of 32 possibilities. The most disorderly possibilities are 3 heads and 2 tails and its reverse. (They are the least structured.) The most disorderly possibilities are also the most likely, with 20 out of 32 possibilities for the 3 heads and 2 tails and its reverse. If we start with an orderly array like 5 heads and toss the coins, it is very likely that we will get a less orderly array as a result, since 30 out of the 32 possibilities are less orderly. So even if you start with an orderly state, there is a strong tendency to go from order to disorder, from low entropy to high entropy. The reverse can happen, but it is unlikely.

Macrostate		Number of microstates
Heads	Tails	(W)
100	0	1
99	1	1.0×10^2
95	5	7.5×10^7
90	10	1.7×10^{13}

Macrostate		Number of microstates
75	25	2.4×10^{23}
60	40	1.4×10^{28}
55	45	6.1×10^{28}
51	49	9.9×10^{28}
50	50	1.0×10^{29}
49	51	9.9×10^{28}
45	55	6.1×10^{28}
40	60	1.4×10^{28}
25	75	2.4×10^{23}

Macrostate		Number of microstates
10	90	1.7×10^{13}
5	95	7.5×10^7
1	99	1.0×10^2
0	100	1
		Total: 1.27×10^{30}

100-Coin Toss

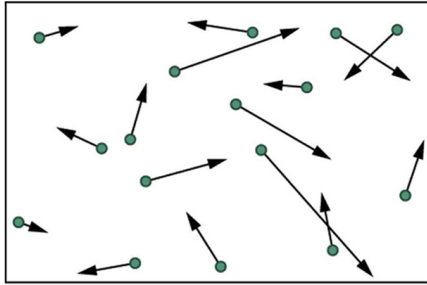
This result becomes dramatic for larger systems. Consider what happens if you have 100 coins instead of just 5. The most orderly arrangements (most structured) are 100 heads or 100 tails. The least orderly (least structured) is that of 50 heads and 50 tails. There is only 1 way (1 microstate) to get the most orderly arrangement of 100 heads. There are 100 ways (100 microstates) to get the next most orderly arrangement of 99 heads and 1 tail (also 100 to get its reverse). And there are 1.0×10^{29} ways to get 50 heads and 50 tails, the least orderly arrangement. [\[link\]](#) is an abbreviated list of the various macrostates and the number of microstates for each macrostate. The total number of microstates—the total number of different ways 100 coins can be tossed—is an impressively large 1.27×10^{30} . Now, if we start with an orderly macrostate like 100 heads and toss the coins, there is a virtual certainty that we will get a less orderly macrostate. If we keep tossing the coins, it is possible, but exceedingly unlikely, that we will ever

get back to the most orderly macrostate. If you tossed the coins once each second, you could expect to get either 100 heads or 100 tails once in 2×10^{22} years! This period is 1 trillion (10^{12}) times longer than the age of the universe, and so the chances are essentially zero. In contrast, there is an 8% chance of getting 50 heads, a 73% chance of getting from 45 to 55 heads, and a 96% chance of getting from 40 to 60 heads. Disorder is highly likely.

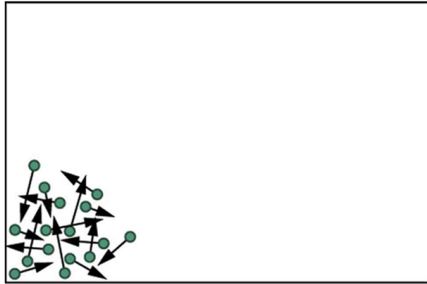
Disorder in a Gas

The fantastic growth in the odds favoring disorder that we see in going from 5 to 100 coins continues as the number of entities in the system increases. Let us now imagine applying this approach to perhaps a small sample of gas. Because counting microstates and macrostates involves statistics, this is called **statistical analysis**. The macrostates of a gas correspond to its macroscopic properties, such as volume, temperature, and pressure; and its microstates correspond to the detailed description of the positions and velocities of its atoms. Even a small amount of gas has a huge number of atoms: 1.0 cm^3 of an ideal gas at 1.0 atm and 0° C has 2.7×10^{19} atoms. So each macrostate has an immense number of microstates. In plain language, this means that there are an immense number of ways in which the atoms in a gas can be arranged, while still having the same pressure, temperature, and so on.

The most likely conditions (or macrostates) for a gas are those we see all the time—a random distribution of atoms in space with a Maxwell-Boltzmann distribution of speeds in random directions, as predicted by kinetic theory. This is the most disorderly and least structured condition we can imagine. In contrast, one type of very orderly and structured macrostate has all of the atoms in one corner of a container with identical velocities. There are very few ways to accomplish this (very few microstates corresponding to it), and so it is exceedingly unlikely ever to occur. (See [\[link\]](#)(b).) Indeed, it is so unlikely that we have a law saying that it is impossible, which has never been observed to be violated—the second law of thermodynamics.



(a) Likely



(b) Highly unlikely

(a) The ordinary state of gas in a container is a disorderly, random distribution of atoms or molecules with a Maxwell-Boltzmann distribution of speeds. It is so unlikely that these atoms or molecules would ever end up in one corner of the container that it might as well be impossible. (b) With energy transfer, the gas can be forced into one corner and its entropy greatly reduced. But left alone, it will

spontaneously
increase its entropy
and return to the
normal conditions,
because they are
immensely more
likely.

The disordered condition is one of high entropy, and the ordered one has low entropy. With a transfer of energy from another system, we could force all of the atoms into one corner and have a local decrease in entropy, but at the cost of an overall increase in entropy of the universe. If the atoms start out in one corner, they will quickly disperse and become uniformly distributed and will never return to the orderly original state ([\[link\]](#)(b)). Entropy will increase. With such a large sample of atoms, it is possible—but unimaginably unlikely—for entropy to decrease. Disorder is vastly more likely than order.

The arguments that disorder and high entropy are the most probable states are quite convincing. The great Austrian physicist Ludwig Boltzmann (1844–1906)—who, along with Maxwell, made so many contributions to kinetic theory—proved that the entropy of a system in a given state (a macrostate) can be written as

Equation:

$$S = k \ln W,$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate. W is proportional to the probability that the macrostate will occur. Thus entropy is directly related to the probability of a state—the more likely the state, the greater its entropy. Boltzmann proved that this expression for S is equivalent to the definition $\Delta S = Q/T$, which we have used extensively.

Thus the second law of thermodynamics is explained on a very basic level: entropy either remains the same or increases in every process. This phenomenon is due to the extraordinarily small probability of a decrease, based on the extraordinarily larger number of microstates in systems with greater entropy. Entropy *can* decrease, but for any macroscopic system, this outcome is so unlikely that it will never be observed.

Example:**Entropy Increases in a Coin Toss**

Suppose you toss 100 coins starting with 60 heads and 40 tails, and you get the most likely result, 50 heads and 50 tails. What is the change in entropy?

Strategy

Noting that the number of microstates is labeled W in [\[link\]](#) for the 100-coin toss, we can use $\Delta S = S_f - S_i = k \ln W_f - k \ln W_i$ to calculate the change in entropy.

Solution

The change in entropy is

Equation:

$$\Delta S = S_f - S_i = k \ln W_f - k \ln W_i,$$

where the subscript i stands for the initial 60 heads and 40 tails state, and the subscript f for the final 50 heads and 50 tails state. Substituting the values for W from [\[link\]](#) gives

Equation:

$$\begin{aligned}\Delta S &= (1.38 \times 10^{-23} \text{ J/K}) [\ln(1.0 \times 10^{29}) - \ln(1.4 \times 10^{28})] \\ &= 2.7 \times 10^{-23} \text{ J/K}\end{aligned}$$

Discussion

This increase in entropy means we have moved to a less orderly situation. It is not impossible for further tosses to produce the initial state of 60 heads and 40 tails, but it is less likely. There is about a 1 in 90 chance for that decrease in entropy ($-2.7 \times 10^{-23} \text{ J/K}$) to occur. If we calculate the decrease in entropy to move to the most orderly state, we get

$\Delta S = -92 \times 10^{-23} \text{ J/K}$. There is about a 1 in 10^{30} chance of this change occurring. So while very small decreases in entropy are unlikely, slightly greater decreases are impossibly unlikely. These probabilities imply, again, that for a macroscopic system, a decrease in entropy is impossible. For example, for heat transfer to occur spontaneously from 1.00 kg of 0°C ice to its 0°C environment, there would be a decrease in entropy of $1.22 \times 10^3 \text{ J/K}$. Given that a ΔS of 10^{-21} J/K corresponds to about a 1 in 10^{30} chance, a decrease of this size (10^3 J/K) is an *utter* impossibility. Even for a milligram of melted ice to spontaneously refreeze is impossible.

Note:

Problem-Solving Strategies for Entropy

1. *Examine the situation to determine if entropy is involved.*
2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns). You must carefully identify the heat transfer, if any, and the temperature at which the process takes place. It is also important to identify the initial and final states.*
5. *Solve the appropriate equation for the quantity to be determined (the unknown). Note that the change in entropy can be determined between any states by calculating it for a reversible process.*
6. *Substitute the known value along with their units into the appropriate equation, and obtain numerical solutions complete with units.*
7. *To see if it is reasonable: Does it make sense?* For example, total entropy should increase for any real process or be constant for a reversible process. Disordered states should be more probable and have greater entropy than ordered states.

Section Summary

- Disorder is far more likely than order, which can be seen statistically.
- The entropy of a system in a given state (a macrostate) can be written

as

Equation:

$$S = k \ln W,$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate.

Conceptual Questions

Exercise:

Problem:

Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problem Exercises

Exercise:

Problem:

Using [\[link\]](#), verify the contention that if you toss 100 coins each second, you can expect to get 100 heads or 100 tails once in 2×10^{22} years; calculate the time to two-digit accuracy.

Solution:

It should happen twice in every 1.27×10^{30} s or once in every

$$6.35 \times 10^{29} \text{ s} \left(6.35 \times 10^{29} \text{ s} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) \left(\frac{1 \text{ d}}{24 \text{ h}} \right) \left(\frac{1 \text{ y}}{365.25 \text{ d}} \right)$$

$$= 2.0 \times 10^{22} \text{ y}$$

Exercise:

Problem:

What percent of the time will you get something in the range from 60 heads and 40 tails through 40 heads and 60 tails when tossing 100 coins? The total number of microstates in that range is 1.22×10^{30} . (Consult [\[link\]](#).)

Exercise:

Problem:

(a) If tossing 100 coins, how many ways (microstates) are there to get the three most likely macrostates of 49 heads and 51 tails, 50 heads and 50 tails, and 51 heads and 49 tails? (b) What percent of the total possibilities is this? (Consult [\[link\]](#).)

Solution:

(a) 3.0×10^{29}

(b) 24%

Exercise:

Problem:

(a) What is the change in entropy if you start with 100 coins in the 45 heads and 55 tails macrostate, toss them, and get 51 heads and 49 tails? (b) What if you get 75 heads and 25 tails? (c) How much more likely is 51 heads and 49 tails than 75 heads and 25 tails? (d) Does either outcome violate the second law of thermodynamics?

Exercise:

Problem:

(a) What is the change in entropy if you start with 10 coins in the 5 heads and 5 tails macrostate, toss them, and get 2 heads and 8 tails? (b) How much more likely is 5 heads and 5 tails than 2 heads and 8 tails? (Take the ratio of the number of microstates to find out.) (c) If you were betting on 2 heads and 8 tails would you accept odds of 252 to 45? Explain why or why not.

Solution:

(a) $-2.38 \times 10^{-23} \text{ J/K}$

(b) 5.6 times more likely

(c) If you were betting on two heads and 8 tails, the odds of breaking even are 252 to 45, so on average you would break even. So, no, you wouldn't bet on odds of 252 to 45.

Macrostate		Number of Microstates (W)
Heads	Tails	
10	0	1
9	1	10
8	2	45
7	3	120

Macrostate		Number of Microstates (W)
6	4	210
5	5	252
4	6	210
3	7	120
2	8	45
1	9	10
0	10	1
		Total: 1024

10-Coin Toss

Exercise:

Problem:

(a) If you toss 10 coins, what percent of the time will you get the three most likely macrostates (6 heads and 4 tails, 5 heads and 5 tails, 4 heads and 6 tails)? (b) You can realistically toss 10 coins and count the number of heads and tails about twice a minute. At that rate, how long will it take on average to get either 10 heads and 0 tails or 0 heads and 10 tails?

Exercise:

Problem:

(a) Construct a table showing the macrostates and all of the individual microstates for tossing 6 coins. (Use [\[link\]](#) as a guide.) (b) How many macrostates are there? (c) What is the total number of microstates? (d) What percent chance is there of tossing 5 heads and 1 tail? (e) How much more likely are you to toss 3 heads and 3 tails than 5 heads and 1 tail? (Take the ratio of the number of microstates to find out.)

Solution:

(b) 7

(c) 64

(d) 9.38%

(e) 3.33 times more likely (20 to 6)

Exercise:**Problem:**

In an air conditioner, 12.65 MJ of heat transfer occurs from a cold environment in 1.00 h. (a) What mass of ice melting would involve the same heat transfer? (b) How many hours of operation would be equivalent to melting 900 kg of ice? (c) If ice costs 20 cents per kg, do you think the air conditioner could be operated more cheaply than by simply using ice? Describe in detail how you evaluate the relative costs.

Glossary

macrostate

an overall property of a system

microstate

each sequence within a larger macrostate

statistical analysis

using statistics to examine data, such as counting microstates and macrostates

Introduction to Oscillatory Motion and Waves

class="introduction"

There
are at
least
four
types
of
waves
in this
picture
—only
the
water
waves
are
evident
. There
are also
sound
waves,
light
waves,
and
waves
on the
guitar
strings.
(credit:
John
Norton
)



What do an ocean buoy, a child in a swing, the cone inside a speaker, a guitar, atoms in a crystal, the motion of chest cavities, and the beating of hearts all have in common? They all **oscillate**—that is, they move back and forth between two points. Many systems oscillate, and they have certain characteristics in common. All oscillations involve force and energy. You push a child in a swing to get the motion started. The energy of atoms vibrating in a crystal can be increased with heat. You put energy into a guitar string when you pluck it.

Some oscillations create **waves**. A guitar creates sound waves. You can make water waves in a swimming pool by slapping the water with your hand. You can no doubt think of other types of waves. Some, such as water waves, are visible. Some, such as sound waves, are not. But *every wave is a disturbance that moves from its source and carries energy*. Other examples of waves include earthquakes and visible light. Even subatomic particles, such as electrons, can behave like waves.

By studying oscillatory motion and waves, we shall find that a small number of underlying principles describe all of them and that wave phenomena are more common than you have ever imagined. We begin by studying the type of force that underlies the simplest oscillations and waves. We will then expand our exploration of oscillatory motion and waves to

include concepts such as simple harmonic motion, uniform circular motion, and damped harmonic motion. Finally, we will explore what happens when two or more waves share the same space, in the phenomena known as superposition and interference.

Glossary

oscillate

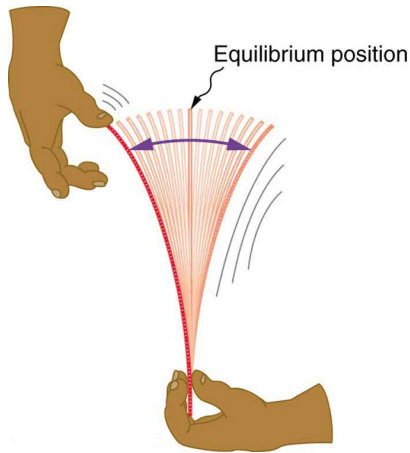
moving back and forth regularly between two points

wave

a disturbance that moves from its source and carries energy

Hooke's Law: Stress and Strain Revisited

- Explain Newton's third law of motion with respect to stress and deformation.
- Describe the restoration of force and displacement.
- Calculate the energy in Hooke's Law of deformation, and the stored energy in a spring.



When displaced from its vertical equilibrium position, this plastic ruler oscillates back and forth because of the restoring force opposing displacement.

When the ruler is on the left, there is a force to the right, and vice versa.

Newton's first law implies that an object oscillating back and forth is experiencing forces. Without force, the object would move in a straight line

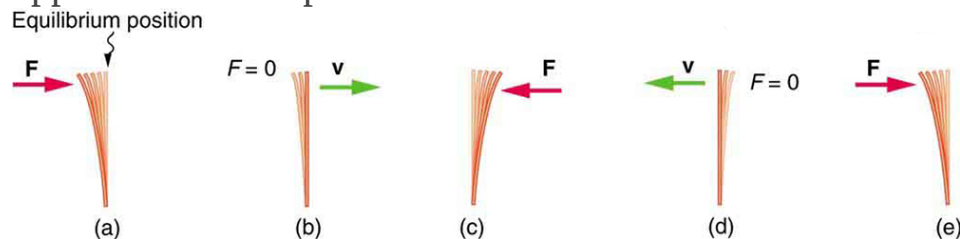
at a constant speed rather than oscillate. Consider, for example, plucking a plastic ruler to the left as shown in [\[link\]](#). The deformation of the ruler creates a force in the opposite direction, known as a **restoring force**. Once released, the restoring force causes the ruler to move back toward its stable equilibrium position, where the net force on it is zero. However, by the time the ruler gets there, it gains momentum and continues to move to the right, producing the opposite deformation. It is then forced to the left, back through equilibrium, and the process is repeated until dissipative forces dampen the motion. These forces remove mechanical energy from the system, gradually reducing the motion until the ruler comes to rest.

The simplest oscillations occur when the restoring force is directly proportional to displacement. When stress and strain were covered in [Newton's Third Law of Motion](#), the name was given to this relationship between force and displacement was Hooke's law:

Equation:

$$F = -kx.$$

Here, F is the restoring force, x is the displacement from equilibrium or **deformation**, and k is a constant related to the difficulty in deforming the system. The minus sign indicates the restoring force is in the direction opposite to the displacement.

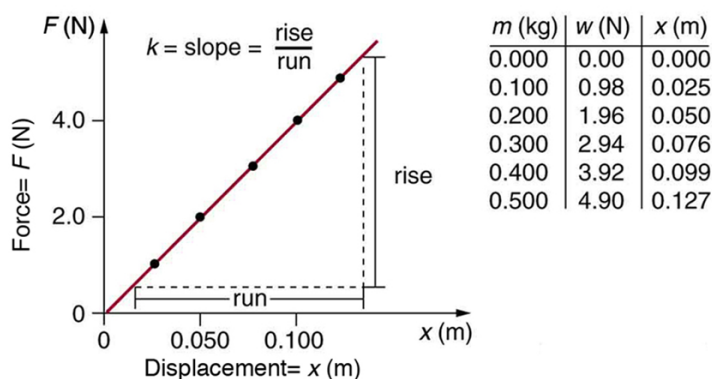


(a) The plastic ruler has been released, and the restoring force is returning the ruler to its equilibrium position. (b) The net force is zero at the equilibrium position, but the ruler has momentum and continues to move to the right. (c) The restoring force is in the opposite direction. It stops the ruler and moves it back toward equilibrium again. (d) Now the ruler has momentum to the left. (e) In the absence of damping

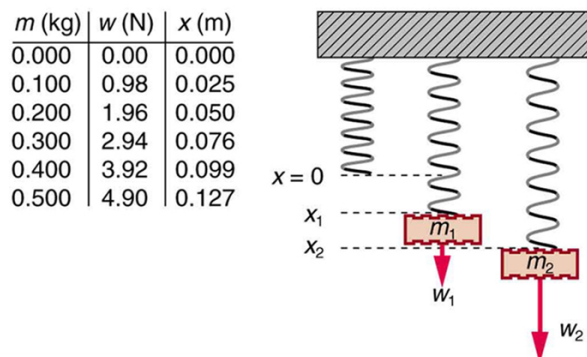
(caused by frictional forces), the ruler reaches its original position. From there, the motion will repeat itself.

The **force constant** k is related to the rigidity (or stiffness) of a system—the larger the force constant, the greater the restoring force, and the stiffer the system. The units of k are newtons per meter (N/m). For example, k is directly related to Young’s modulus when we stretch a string. [\[link\]](#) shows a graph of the absolute value of the restoring force versus the displacement for a system that can be described by Hooke’s law—a simple spring in this case. The slope of the graph equals the force constant k in newtons per meter. A common physics laboratory exercise is to measure restoring forces created by springs, determine if they follow Hooke’s law, and calculate their force constants if they do.

a)



b)



(a) A graph of absolute value of the restoring force versus displacement is

displayed. The fact that the graph is a straight line means that the system obeys Hooke's law. The slope of the graph is the force constant k . (b) The data in the graph were generated by measuring the displacement of a spring from equilibrium while supporting various weights. The restoring force equals the weight supported, if the mass is stationary.

Example:
How Stiff Are Car Springs?



The mass of a car increases due to the introduction of a passenger. This affects the displacement of

the car on its
suspension
system. (credit:
exfordy on
Flickr)

What is the force constant for the suspension system of a car that settles 1.20 cm when an 80.0-kg person gets in?

Strategy

Consider the car to be in its equilibrium position $x = 0$ before the person gets in. The car then settles down 1.20 cm, which means it is displaced to a position $x = -1.20 \times 10^{-2}$ m. At that point, the springs supply a restoring force F equal to the person's weight

$w = mg = (80.0 \text{ kg})(9.80 \text{ m/s}^2) = 784 \text{ N}$. We take this force to be F in Hooke's law. Knowing F and x , we can then solve the force constant k .

Solution

1. Solve Hooke's law, $F = -kx$, for k :

Equation:

$$k = -\frac{F}{x}.$$

Substitute known values and solve k :

Equation:

$$\begin{aligned} k &= -\frac{784 \text{ N}}{-1.20 \times 10^{-2} \text{ m}} \\ &= 6.53 \times 10^4 \text{ N/m.} \end{aligned}$$

Discussion

Note that F and x have opposite signs because they are in opposite directions—the restoring force is up, and the displacement is down. Also, note that the car would oscillate up and down when the person got in if it

were not for damping (due to frictional forces) provided by shock absorbers. Bouncing cars are a sure sign of bad shock absorbers.

Energy in Hooke's Law of Deformation

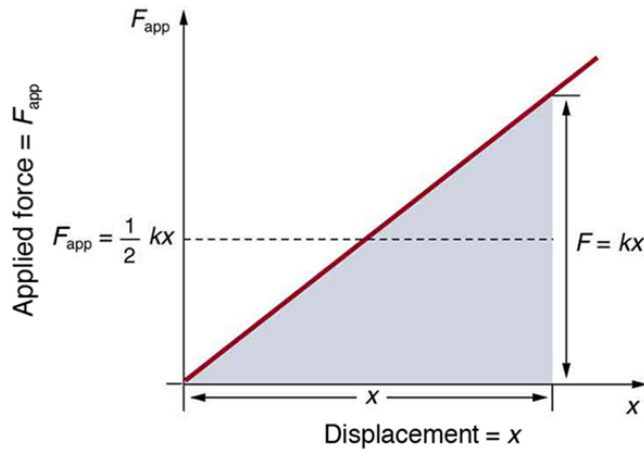
In order to produce a deformation, work must be done. That is, a force must be exerted through a distance, whether you pluck a guitar string or compress a car spring. If the only result is deformation, and no work goes into thermal, sound, or kinetic energy, then all the work is initially stored in the deformed object as some form of potential energy. The potential energy stored in a spring is $PE_{el} = \frac{1}{2}kx^2$. Here, we generalize the idea to elastic potential energy for a deformation of any system that can be described by Hooke's law. Hence,

Equation:

$$PE_{el} = \frac{1}{2}kx^2,$$

where PE_{el} is the **elastic potential energy** stored in any deformed system that obeys Hooke's law and has a displacement x from equilibrium and a force constant k .

It is possible to find the work done in deforming a system in order to find the energy stored. This work is performed by an applied force F_{app} . The applied force is exactly opposite to the restoring force (action-reaction), and so $F_{app} = kx$. [\[link\]](#) shows a graph of the applied force versus deformation x for a system that can be described by Hooke's law. Work done on the system is force multiplied by distance, which equals the area under the curve or $(1/2)kx^2$ (Method A in the figure). Another way to determine the work is to note that the force increases linearly from 0 to kx , so that the average force is $(1/2)kx$, the distance moved is x , and thus $W = F_{app}d = [(1/2)kx](x) = (1/2)kx^2$ (Method B in the figure).



Method A

$$W = \frac{1}{2} bh = \frac{1}{2} kxx$$

$$W = \frac{1}{2} kx^2$$

Method B

$$W = f \cdot x = \left(\frac{1}{2} kx \right) (x)$$

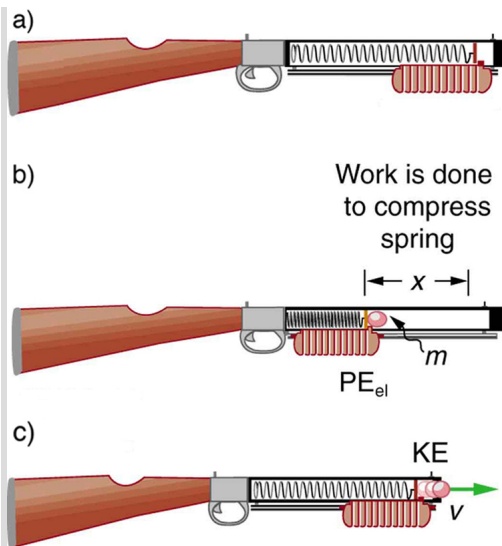
$$W = \frac{1}{2} kx^2$$

A graph of applied force versus distance for the deformation of a system that can be described by Hooke's law is displayed. The work done on the system equals the area under the graph or the area of the triangle, which is half its base multiplied by its height, or $W = (1/2)kx^2$.

Example:

Calculating Stored Energy: A Tranquilizer Gun Spring

We can use a toy gun's spring mechanism to ask and answer two simple questions: (a) How much energy is stored in the spring of a tranquilizer gun that has a force constant of 50.0 N/m and is compressed 0.150 m? (b) If you neglect friction and the mass of the spring, at what speed will a 2.00-g projectile be ejected from the gun?



(a) In this image of the gun, the spring is uncompressed before being cocked. (b) The spring has been compressed a distance x , and the projectile is in place. (c) When released, the spring converts elastic potential energy PE_{el} into kinetic energy.

Strategy for a

(a): The energy stored in the spring can be found directly from elastic potential energy equation, because k and x are given.

Solution for a

Entering the given values for k and x yields

Equation:

$$\begin{aligned} PE_{el} &= \frac{1}{2} kx^2 = \frac{1}{2} (50.0 \text{ N/m})(0.150 \text{ m})^2 = 0.563 \text{ N} \cdot \text{m} \\ &= 0.563 \text{ J} \end{aligned}$$

Strategy for b

Because there is no friction, the potential energy is converted entirely into kinetic energy. The expression for kinetic energy can be solved for the projectile's speed.

Solution for b

1. Identify known quantities:

Equation:

$$KE_f = PE_{el} \text{ or } \frac{1}{2}mv^2 = (1/2)kx^2 = PE_{el} = 0.563 \text{ J}$$

2. Solve for v :

Equation:

$$v = \left[\frac{2PE_{el}}{m} \right]^{1/2} = \left[\frac{2(0.563 \text{ J})}{0.002 \text{ kg}} \right]^{1/2} = 23.7(\text{J/kg})^{1/2}$$

3. Convert units: 23.7 m/s

Discussion

(a) and (b): This projectile speed is impressive for a tranquilizer gun (more than 80 km/h). The numbers in this problem seem reasonable. The force needed to compress the spring is small enough for an adult to manage, and the energy imparted to the dart is small enough to limit the damage it might do. Yet, the speed of the dart is great enough for it to travel an acceptable distance.

Exercise:

Check your Understanding

Problem:

Envision holding the end of a ruler with one hand and deforming it with the other. When you let go, you can see the oscillations of the ruler. In what way could you modify this simple experiment to increase the rigidity of the system?

Solution:

Answer

You could hold the ruler at its midpoint so that the part of the ruler that oscillates is half as long as in the original experiment.

Exercise:

Check your Understanding

Problem:

If you apply a deforming force on an object and let it come to equilibrium, what happened to the work you did on the system?

Solution:

Answer

It was stored in the object as potential energy.

Section Summary

- An oscillation is a back and forth motion of an object between two points of deformation.
- An oscillation may create a wave, which is a disturbance that propagates from where it was created.
- The simplest type of oscillations and waves are related to systems that can be described by Hooke's law:

Equation:

$$F = -kx,$$

where F is the restoring force, x is the displacement from equilibrium or deformation, and k is the force constant of the system.

- Elastic potential energy PE_{el} stored in the deformation of a system that can be described by Hooke's law is given by

Equation:

$$PE_{\text{el}} = (1/2)kx^2.$$

Conceptual Questions

Exercise:

Problem:

Describe a system in which elastic potential energy is stored.

Problems & Exercises

Exercise:

Problem:

Fish are hung on a spring scale to determine their mass (most fishermen feel no obligation to truthfully report the mass).

- (a) What is the force constant of the spring in such a scale if it the spring stretches 8.00 cm for a 10.0 kg load?
- (b) What is the mass of a fish that stretches the spring 5.50 cm?
- (c) How far apart are the half-kilogram marks on the scale?

Solution:

- (a) $1.23 \times 10^3 \text{ N/m}$
- (b) 6.88 kg
- (c) 4.00 mm

Exercise:

Problem:

It is weigh-in time for the local under-85-kg rugby team. The bathroom scale used to assess eligibility can be described by Hooke's law and is depressed 0.75 cm by its maximum load of 120 kg. (a) What is the spring's effective spring constant? (b) A player stands on the scales and depresses it by 0.48 cm. Is he eligible to play on this under-85 kg team?

Exercise:**Problem:**

One type of BB gun uses a spring-driven plunger to blow the BB from its barrel. (a) Calculate the force constant of its plunger's spring if you must compress it 0.150 m to drive the 0.0500-kg plunger to a top speed of 20.0 m/s. (b) What force must be exerted to compress the spring?

Solution:

(a) 889 N/m

(b) 133 N

Exercise:**Problem:**

(a) The springs of a pickup truck act like a single spring with a force constant of 1.30×10^5 N/m. By how much will the truck be depressed by its maximum load of 1000 kg?

(b) If the pickup truck has four identical springs, what is the force constant of each?

Exercise:**Problem:**

When an 80.0-kg man stands on a pogo stick, the spring is compressed 0.120 m.

(a) What is the force constant of the spring? (b) Will the spring be compressed more when he hops down the road?

Solution:

(a) $6.53 \times 10^3 \text{ N/m}$

(b) Yes

Exercise:

Problem:

A spring has a length of 0.200 m when a 0.300-kg mass hangs from it, and a length of 0.750 m when a 1.95-kg mass hangs from it. (a) What is the force constant of the spring? (b) What is the unloaded length of the spring?

Glossary

deformation

displacement from equilibrium

elastic potential energy

potential energy stored as a result of deformation of an elastic object, such as the stretching of a spring

force constant

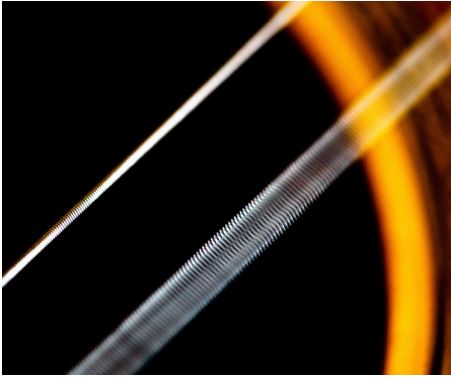
a constant related to the rigidity of a system: the larger the force constant, the more rigid the system; the force constant is represented by k

restoring force

force acting in opposition to the force caused by a deformation

Period and Frequency in Oscillations

- Observe the vibrations of a guitar string.
- Determine the frequency of oscillations.



The strings on this
guitar vibrate at
regular time intervals.
(credit: JAR)

When you pluck a guitar string, the resulting sound has a steady tone and lasts a long time. Each successive vibration of the string takes the same time as the previous one. We define **periodic motion** to be a motion that repeats itself at regular time intervals, such as exhibited by the guitar string or by an object on a spring moving up and down. The time to complete one oscillation remains constant and is called the **period T** . Its units are usually seconds, but may be any convenient unit of time. The word period refers to the time for some event whether repetitive or not; but we shall be primarily interested in periodic motion, which is by definition repetitive. A concept closely related to period is the frequency of an event. For example, if you get a paycheck twice a month, the frequency of payment is two per month and the period between checks is half a month. **Frequency f** is defined to be the number of events per unit time. For periodic motion, frequency is the number of oscillations per unit time. The relationship between frequency and period is

Equation:

$$f = \frac{1}{T}.$$

The SI unit for frequency is the *cycle per second*, which is defined to be a *hertz* (Hz):

Equation:

$$1 \text{ Hz} = 1 \frac{\text{cycle}}{\text{sec}} \text{ or } 1 \text{ Hz} = \frac{1}{\text{s}}$$

A cycle is one complete oscillation. Note that a vibration can be a single or multiple event, whereas oscillations are usually repetitive for a significant number of cycles.

Example:

Determine the Frequency of Two Oscillations: Medical Ultrasound and the Period of Middle C

We can use the formulas presented in this module to determine both the frequency based on known oscillations and the oscillation based on a known frequency. Let's try one example of each. (a) A medical imaging device produces ultrasound by oscillating with a period of $0.400 \mu\text{s}$. What is the frequency of this oscillation? (b) The frequency of middle C on a typical musical instrument is 264 Hz. What is the time for one complete oscillation?

Strategy

Both questions (a) and (b) can be answered using the relationship between period and frequency. In question (a), the period T is given and we are asked to find frequency f . In question (b), the frequency f is given and we are asked to find the period T .

Solution a

1. Substitute $0.400 \mu\text{s}$ for T in $f = \frac{1}{T}$:

Equation:

$$f = \frac{1}{T} = \frac{1}{0.400 \times 10^{-6} \text{ s}}.$$

Solve to find

Equation:

$$f = 2.50 \times 10^6 \text{ Hz}.$$

Discussion a

The frequency of sound found in (a) is much higher than the highest frequency that humans can hear and, therefore, is called ultrasound. Appropriate oscillations at this frequency generate ultrasound used for noninvasive medical diagnoses, such as observations of a fetus in the womb.

Solution b

1. Identify the known values:

The time for one complete oscillation is the period T :

Equation:

$$f = \frac{1}{T}.$$

2. Solve for T :

Equation:

$$T = \frac{1}{f}.$$

3. Substitute the given value for the frequency into the resulting expression:

Equation:

$$T = \frac{1}{f} = \frac{1}{264 \text{ Hz}} = \frac{1}{264 \text{ cycles/s}} = 3.79 \times 10^{-3} \text{ s} = 3.79 \text{ ms}.$$

Discussion

The period found in (b) is the time per cycle, but this value is often quoted as simply the time in convenient units (ms or milliseconds in this case).

Exercise:

Check your Understanding

Problem:

Identify an event in your life (such as receiving a paycheck) that occurs regularly. Identify both the period and frequency of this event.

Solution:

I visit my parents for dinner every other Sunday. The frequency of my visits is 26 per calendar year. The period is two weeks.

Section Summary

- Periodic motion is a repetitious oscillation.
- The time for one oscillation is the period T .
- The number of oscillations per unit time is the frequency f .
- These quantities are related by

Equation:

$$f = \frac{1}{T}.$$

Problems & Exercises

Exercise:

Problem: What is the period of 60.0 Hz electrical power?

Solution:

16.7 ms

Exercise:**Problem:**

If your heart rate is 150 beats per minute during strenuous exercise, what is the time per beat in units of seconds?

Solution:

0.400 s/beats

Exercise:**Problem:**

Find the frequency of a tuning fork that takes 2.50×10^{-3} s to complete one oscillation.

Solution:

400 Hz

Exercise:**Problem:**

A stroboscope is set to flash every 8.00×10^{-5} s. What is the frequency of the flashes?

Solution:

12,500 Hz

Exercise:

Problem:

A tire has a tread pattern with a crevice every 2.00 cm. Each crevice makes a single vibration as the tire moves. What is the frequency of these vibrations if the car moves at 30.0 m/s?

Solution:

1.50 kHz

Exercise:**Problem: Engineering Application**

Each piston of an engine makes a sharp sound every other revolution of the engine. (a) How fast is a race car going if its eight-cylinder engine emits a sound of frequency 750 Hz, given that the engine makes 2000 revolutions per kilometer? (b) At how many revolutions per minute is the engine rotating?

Solution:

(a) 93.8 m/s

(b) 11.3×10^3 rev/min

Glossary

period

time it takes to complete one oscillation

periodic motion

motion that repeats itself at regular time intervals

frequency

number of events per unit of time

Simple Harmonic Motion: A Special Periodic Motion

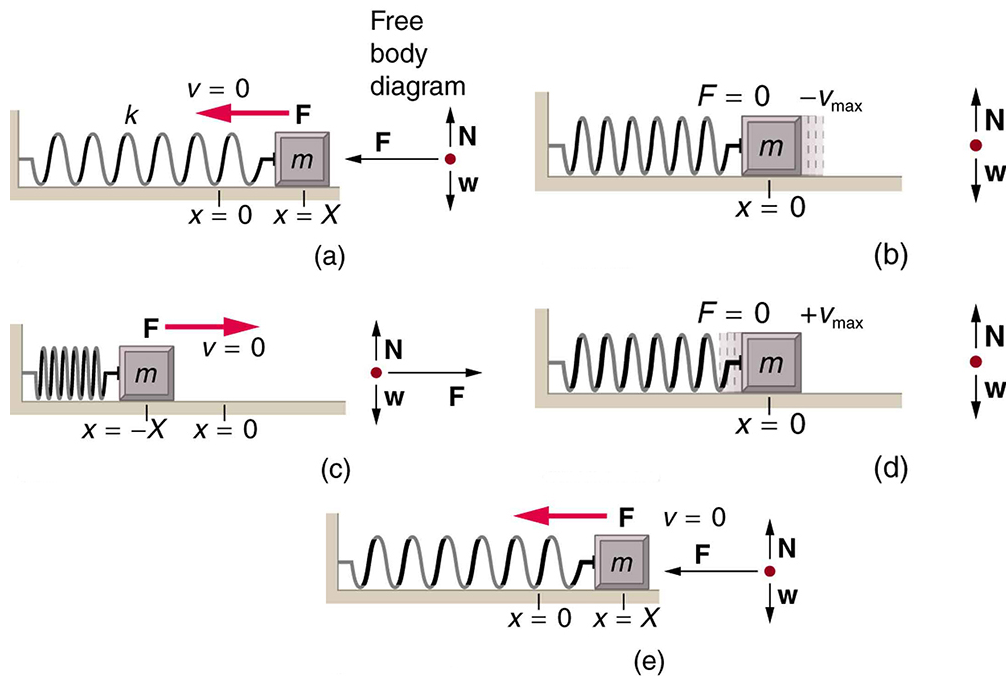
- Describe a simple harmonic oscillator.
- Explain the link between simple harmonic motion and waves.

The oscillations of a system in which the net force can be described by Hooke's law are of special importance, because they are very common. They are also the simplest oscillatory systems. **Simple Harmonic Motion** (SHM) is the name given to oscillatory motion for a system where the net force can be described by Hooke's law, and such a system is called a **simple harmonic oscillator**. If the net force can be described by Hooke's law and there is no *damping* (by friction or other non-conservative forces), then a simple harmonic oscillator will oscillate with equal displacement on either side of the equilibrium position, as shown for an object on a spring in [\[link\]](#). The maximum displacement from equilibrium is called the **amplitude** X . The units for amplitude and displacement are the same, but depend on the type of oscillation. For the object on the spring, the units of amplitude and displacement are meters; whereas for sound oscillations, they have units of pressure (and other types of oscillations have yet other units). Because amplitude is the maximum displacement, it is related to the energy in the oscillation.

Note:

Take-Home Experiment: SHM and the Marble

Find a bowl or basin that is shaped like a hemisphere on the inside. Place a marble inside the bowl and tilt the bowl periodically so the marble rolls from the bottom of the bowl to equally high points on the sides of the bowl. Get a feel for the force required to maintain this periodic motion. What is the restoring force and what role does the force you apply play in the simple harmonic motion (SHM) of the marble?



An object attached to a spring sliding on a frictionless surface is an uncomplicated simple harmonic oscillator. When displaced from equilibrium, the object performs simple harmonic motion that has an amplitude X and a period T . The object's maximum speed occurs as it passes through equilibrium. The stiffer the spring is, the smaller the period T . The greater the mass of the object is, the greater the period T .

What is so significant about simple harmonic motion? One special thing is that the period T and frequency f of a simple harmonic oscillator are independent of amplitude. The string of a guitar, for example, will oscillate with the same frequency whether plucked gently or hard. Because the period is constant, a simple harmonic oscillator can be used as a clock.

Two important factors do affect the period of a simple harmonic oscillator. The period is related to how stiff the system is. A very stiff object has a large force constant k , which causes the system to have a smaller period. For example, you can adjust a diving board's stiffness—the stiffer it is, the

faster it vibrates, and the shorter its period. Period also depends on the mass of the oscillating system. The more massive the system is, the longer the period. For example, a heavy person on a diving board bounces up and down more slowly than a light one.

In fact, the mass m and the force constant k are the *only* factors that affect the period and frequency of simple harmonic motion.

Note:

Period of Simple Harmonic Oscillator

The *period of a simple harmonic oscillator* is given by

Equation:

$$T = 2\pi\sqrt{\frac{m}{k}}$$

and, because $f = 1/T$, the *frequency of a simple harmonic oscillator* is

Equation:

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}.$$

Note that neither T nor f has any dependence on amplitude.

Note:

Take-Home Experiment: Mass and Ruler Oscillations

Find two identical wooden or plastic rulers. Tape one end of each ruler firmly to the edge of a table so that the length of each ruler that protrudes from the table is the same. On the free end of one ruler tape a heavy object such as a few large coins. Pluck the ends of the rulers at the same time and observe which one undergoes more cycles in a time period, and measure the period of oscillation of each of the rulers.

Example:**Calculate the Frequency and Period of Oscillations: Bad Shock Absorbers in a Car**

If the shock absorbers in a car go bad, then the car will oscillate at the least provocation, such as when going over bumps in the road and after stopping (See [\[link\]](#)). Calculate the frequency and period of these oscillations for such a car if the car's mass (including its load) is 900 kg and the force constant (k) of the suspension system is 6.53×10^4 N/m.

Strategy

The frequency of the car's oscillations will be that of a simple harmonic oscillator as given in the equation $f = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$. The mass and the force constant are both given.

Solution

1. Enter the known values of k and m :

Equation:

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}} = \frac{1}{2\pi} \sqrt{\frac{6.53 \times 10^4 \text{ N/m}}{900 \text{ kg}}}.$$

2. Calculate the frequency:

Equation:

$$\frac{1}{2\pi} \sqrt{72.6/\text{s}^{-2}} = 1.3656/\text{s}^{-1} \approx 1.36/\text{s}^{-1} = 1.36 \text{ Hz}.$$

3. You could use $T = 2\pi\sqrt{\frac{m}{k}}$ to calculate the period, but it is simpler to use the relationship $T = 1/f$ and substitute the value just found for f :

Equation:

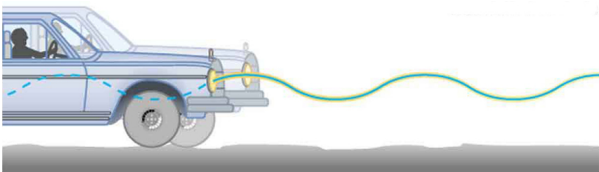
$$T = \frac{1}{f} = \frac{1}{1.356 \text{ Hz}} = 0.738 \text{ s}.$$

Discussion

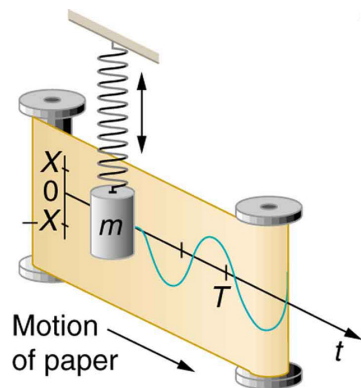
The values of T and f both seem about right for a bouncing car. You can observe these oscillations if you push down hard on the end of a car and let go.

The Link between Simple Harmonic Motion and Waves

If a time-exposure photograph of the bouncing car were taken as it drove by, the headlight would make a wavelike streak, as shown in [\[link\]](#). Similarly, [\[link\]](#) shows an object bouncing on a spring as it leaves a wavelike "trace of its position on a moving strip of paper. Both waves are sine functions. All simple harmonic motion is intimately related to sine and cosine waves.



The bouncing car makes a wavelike motion. If the restoring force in the suspension system can be described only by Hooke's law, then the wave is a sine function. (The wave is the trace produced by the headlight as the car moves to the right.)



The vertical position of an object bouncing on a spring is recorded on a strip of moving paper, leaving a sine wave.

The displacement as a function of time t in any simple harmonic motion—that is, one in which the net restoring force can be described by Hooke's law, is given by

Equation:

$$x(t) = X \cos \frac{2\pi t}{T},$$

where X is amplitude. At $t = 0$, the initial position is $x_0 = X$, and the displacement oscillates back and forth with a period T . (When $t = T$, we get $x = X$ again because $\cos 2\pi = 1$.) Furthermore, from this expression for x , the velocity v as a function of time is given by:

Equation:

$$v(t) = -v_{\max} \sin \left(\frac{2\pi t}{T} \right),$$

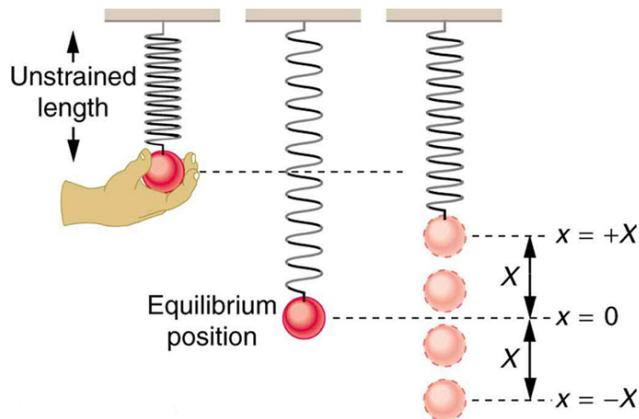
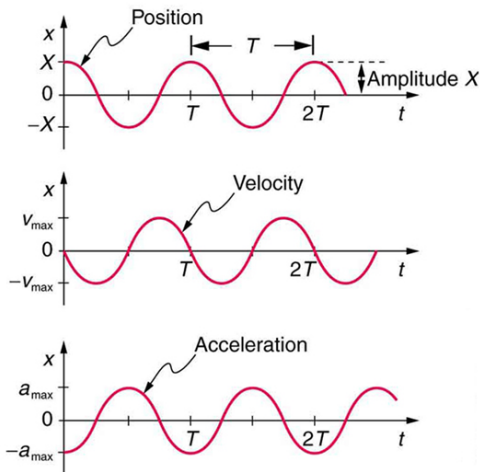
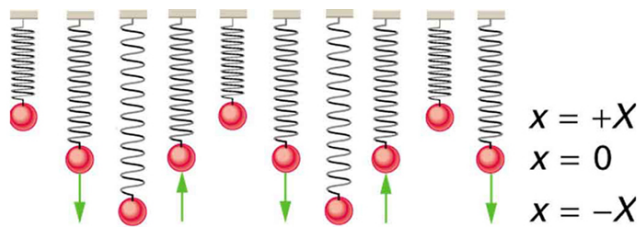
where $v_{\max} = 2\pi X/T = X\sqrt{k/m}$. The object has zero velocity at maximum displacement—for example, $v = 0$ when $t = 0$, and at that time $x = X$. The minus sign in the first equation for $v(t)$ gives the correct direction for the velocity. Just after the start of the motion, for instance, the velocity is negative because the system is moving back toward the equilibrium point. Finally, we can get an expression for acceleration using Newton's second law. [Then we have $x(t)$, $v(t)$, t , and $a(t)$, the quantities needed for kinematics and a description of simple harmonic motion.] According to Newton's second law, the acceleration is $a = F/m = kx/m$. So, $a(t)$ is also a cosine function:

Equation:

$$a(t) = -\frac{kX}{m} \cos \frac{2\pi t}{T}.$$

Hence, $a(t)$ is directly proportional to and in the opposite direction to $x(t)$.

[\[link\]](#) shows the simple harmonic motion of an object on a spring and presents graphs of $x(t)$, $v(t)$, and $a(t)$ versus time.



Graphs of $x(t)$, $v(t)$, and $a(t)$ versus t for the motion of an object on a spring. The net force on the object can be described by Hooke's law, and so the object undergoes simple harmonic motion. Note that the initial position has the vertical displacement at its maximum value X ; v is initially zero and then negative as the object moves down; and the initial acceleration

is negative, back toward the equilibrium position and becomes zero at that point.

The most important point here is that these equations are mathematically straightforward and are valid for all simple harmonic motion. They are very useful in visualizing waves associated with simple harmonic motion, including visualizing how waves add with one another.

Exercise:

Check Your Understanding

Problem:

Suppose you pluck a banjo string. You hear a single note that starts out loud and slowly quiets over time. Describe what happens to the sound waves in terms of period, frequency and amplitude as the sound decreases in volume.

Solution:

Frequency and period remain essentially unchanged. Only amplitude decreases as volume decreases.

Exercise:

Check Your Understanding

Problem:

A babysitter is pushing a child on a swing. At the point where the swing reaches x , where would the corresponding point on a wave of this motion be located?

Solution:

x is the maximum deformation, which corresponds to the amplitude of the wave. The point on the wave would either be at the very top or the very bottom of the curve.

Note:**PhET Explorations: Masses and Springs**

A realistic mass and spring laboratory. Hang masses from springs and adjust the spring stiffness and damping. You can even slow time. Transport the lab to different planets. A chart shows the kinetic, potential, and thermal energy for each spring.

https://phet.colorado.edu/sims/mass-spring-lab/mass-spring-lab_en.html

Section Summary

- Simple harmonic motion is oscillatory motion for a system that can be described only by Hooke's law. Such a system is also called a simple harmonic oscillator.
- Maximum displacement is the amplitude X . The period T and frequency f of a simple harmonic oscillator are given by

$T = 2\pi\sqrt{\frac{m}{k}}$ and $f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$, where m is the mass of the system.

- Displacement in simple harmonic motion as a function of time is given by $x(t) = X \cos \frac{2\pi t}{T}$.
- The velocity is given by $v(t) = -v_{\max} \sin \frac{2\pi t}{T}$, where $v_{\max} = \sqrt{k/m}X$.
- The acceleration is found to be $a(t) = -\frac{kX}{m} \cos \frac{2\pi t}{T}$.

Conceptual Questions

Exercise:**Problem:**

What conditions must be met to produce simple harmonic motion?

Exercise:

Problem:

- (a) If frequency is not constant for some oscillation, can the oscillation be simple harmonic motion?
- (b) Can you think of any examples of harmonic motion where the frequency may depend on the amplitude?

Exercise:**Problem:**

Give an example of a simple harmonic oscillator, specifically noting how its frequency is independent of amplitude.

Exercise:**Problem:**

Explain why you expect an object made of a stiff material to vibrate at a higher frequency than a similar object made of a spongy material.

Exercise:**Problem:**

As you pass a freight truck with a trailer on a highway, you notice that its trailer is bouncing up and down slowly. Is it more likely that the trailer is heavily loaded or nearly empty? Explain your answer.

Exercise:**Problem:**

Some people modify cars to be much closer to the ground than when manufactured. Should they install stiffer springs? Explain your answer.

Problems & Exercises**Exercise:**

Problem:

A type of cuckoo clock keeps time by having a mass bouncing on a spring, usually something cute like a cherub in a chair. What force constant is needed to produce a period of 0.500 s for a 0.0150-kg mass?

Solution:

2.37 N/m

Exercise:**Problem:**

If the spring constant of a simple harmonic oscillator is doubled, by what factor will the mass of the system need to change in order for the frequency of the motion to remain the same?

Exercise:**Problem:**

A 0.500-kg mass suspended from a spring oscillates with a period of 1.50 s. How much mass must be added to the object to change the period to 2.00 s?

Solution:

0.389 kg

Exercise:**Problem:**

By how much leeway (both percentage and mass) would you have in the selection of the mass of the object in the previous problem if you did not wish the new period to be greater than 2.01 s or less than 1.99 s?

Exercise:

Problem:

Suppose you attach the object with mass m to a vertical spring originally at rest, and let it bounce up and down. You release the object from rest at the spring's original rest length. (a) Show that the spring exerts an upward force of $2.00\ mg$ on the object at its lowest point. (b) If the spring has a force constant of $10.0\ \text{N/m}$ and a 0.25-kg -mass object is set in motion as described, find the amplitude of the oscillations. (c) Find the maximum velocity.

Exercise:**Problem:**

A diver on a diving board is undergoing simple harmonic motion. Her mass is $55.0\ \text{kg}$ and the period of her motion is $0.800\ \text{s}$. The next diver is a male whose period of simple harmonic oscillation is $1.05\ \text{s}$. What is his mass if the mass of the board is negligible?

Solution:

$94.7\ \text{kg}$

Exercise:**Problem:**

Suppose a diving board with no one on it bounces up and down in a simple harmonic motion with a frequency of $4.00\ \text{Hz}$. The board has an effective mass of $10.0\ \text{kg}$. What is the frequency of the simple harmonic motion of a 75.0-kg diver on the board?

Exercise:**Problem:**



This child's toy
relies on springs to
keep infants
entertained. (credit:
By Humboldtthead,
Flickr)

The device pictured in [\[link\]](#) entertains infants while keeping them from wandering. The child bounces in a harness suspended from a door frame by a spring constant.

(a) If the spring stretches 0.250 m while supporting an 8.0-kg child, what is its spring constant?

(b) What is the time for one complete bounce of this child? (c) What is the child's maximum velocity if the amplitude of her bounce is 0.200 m?

Exercise:

Problem:

A 90.0-kg skydiver hanging from a parachute bounces up and down with a period of 1.50 s. What is the new period of oscillation when a second skydiver, whose mass is 60.0 kg, hangs from the legs of the first, as seen in [\[link\]](#).



The oscillations of one skydiver are about to be affected by a second skydiver. (credit: U.S. Army, www.army.mil)

Solution:

1.94 s

Glossary

amplitude

the maximum displacement from the equilibrium position of an object oscillating around the equilibrium position

simple harmonic motion

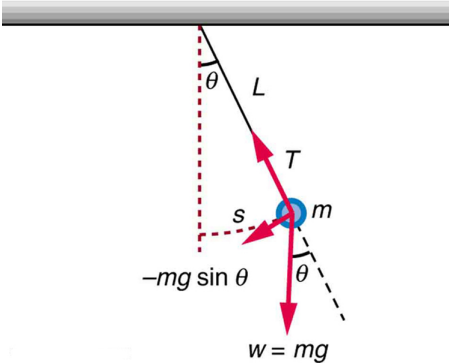
the oscillatory motion in a system where the net force can be described by Hooke's law

simple harmonic oscillator

a device that implements Hooke's law, such as a mass that is attached to a spring, with the other end of the spring being connected to a rigid support such as a wall

The Simple Pendulum

- Measure acceleration due to gravity.



A simple pendulum has a small-diameter bob and a string that has a very small mass but is strong enough not to stretch appreciably. The linear displacement from equilibrium is s , the length of the arc. Also shown are the forces on the bob, which result in a net force of $-mg \sin \theta$ toward the equilibrium position—that is, a restoring force.

Pendulums are in common usage. Some have crucial uses, such as in clocks; some are for fun, such as a child's swing; and some are just there, such as the sinker on a fishing line. For small displacements, a pendulum is a simple harmonic oscillator. A **simple pendulum** is defined to have an

object that has a small mass, also known as the pendulum bob, which is suspended from a light wire or string, such as shown in [\[link\]](#). Exploring the simple pendulum a bit further, we can discover the conditions under which it performs simple harmonic motion, and we can derive an interesting expression for its period.

We begin by defining the displacement to be the arc length s . We see from [\[link\]](#) that the net force on the bob is tangent to the arc and equals $-mg \sin \theta$. (The weight mg has components $mg \cos \theta$ along the string and $mg \sin \theta$ tangent to the arc.) Tension in the string exactly cancels the component $mg \cos \theta$ parallel to the string. This leaves a *net* restoring force back toward the equilibrium position at $\theta = 0$.

Now, if we can show that the restoring force is directly proportional to the displacement, then we have a simple harmonic oscillator. In trying to determine if we have a simple harmonic oscillator, we should note that for small angles (less than about 15°), $\sin \theta \approx \theta$ ($\sin \theta$ and θ differ by about 1% or less at smaller angles). Thus, for angles less than about 15° , the restoring force F is

Equation:

$$F \approx -mg\theta.$$

The displacement s is directly proportional to θ . When θ is expressed in radians, the arc length in a circle is related to its radius (L in this instance) by:

Equation:

$$s = L\theta,$$

so that

Equation:

$$\theta = \frac{s}{L}.$$

For small angles, then, the expression for the restoring force is:

Equation:

$$F \approx -\frac{mg}{L}s$$

This expression is of the form:

Equation:

$$F = -kx,$$

where the force constant is given by $k = mg/L$ and the displacement is given by $x = s$. For angles less than about 15° , the restoring force is directly proportional to the displacement, and the simple pendulum is a simple harmonic oscillator.

Using this equation, we can find the period of a pendulum for amplitudes less than about 15° . For the simple pendulum:

Equation:

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{m}{mg/L}}.$$

Thus,

Equation:

$$T = 2\pi\sqrt{\frac{L}{g}}$$

for the period of a simple pendulum. This result is interesting because of its simplicity. The only things that affect the period of a simple pendulum are its length and the acceleration due to gravity. The period is completely independent of other factors, such as mass. As with simple harmonic oscillators, the period T for a pendulum is nearly independent of amplitude,

especially if θ is less than about 15° . Even simple pendulum clocks can be finely adjusted and accurate.

Note the dependence of T on g . If the length of a pendulum is precisely known, it can actually be used to measure the acceleration due to gravity. Consider the following example.

Example:

Measuring Acceleration due to Gravity: The Period of a Pendulum

What is the acceleration due to gravity in a region where a simple pendulum having a length 75.000 cm has a period of 1.7357 s?

Strategy

We are asked to find g given the period T and the length L of a pendulum.

We can solve $T = 2\pi\sqrt{\frac{L}{g}}$ for g , assuming only that the angle of deflection is less than 15° .

Solution

1. Square $T = 2\pi\sqrt{\frac{L}{g}}$ and solve for g :

Equation:

$$g = 4\pi^2 \frac{L}{T^2}.$$

2. Substitute known values into the new equation:

Equation:

$$g = 4\pi^2 \frac{0.75000 \text{ m}}{(1.7357 \text{ s})^2}.$$

3. Calculate to find g :

Equation:

$$g = 9.8281 \text{ m/s}^2.$$

Discussion

This method for determining g can be very accurate. This is why length and period are given to five digits in this example. For the precision of the approximation $\sin \theta \approx \theta$ to be better than the precision of the pendulum length and period, the maximum displacement angle should be kept below about 0.5° .

Note:**Making Career Connections**

Knowing g can be important in geological exploration; for example, a map of g over large geographical regions aids the study of plate tectonics and helps in the search for oil fields and large mineral deposits.

Note:**Take Home Experiment: Determining g**

Use a simple pendulum to determine the acceleration due to gravity g in your own locale. Cut a piece of a string or dental floss so that it is about 1 m long. Attach a small object of high density to the end of the string (for example, a metal nut or a car key). Starting at an angle of less than 10° , allow the pendulum to swing and measure the pendulum's period for 10 oscillations using a stopwatch. Calculate g . How accurate is this measurement? How might it be improved?

Exercise:**Check Your Understanding**

Problem:

An engineer builds two simple pendula. Both are suspended from small wires secured to the ceiling of a room. Each pendulum hovers 2 cm above the floor. Pendulum 1 has a bob with a mass of 10 kg. Pendulum 2 has a bob with a mass of 100 kg. Describe how the motion of the pendula will differ if the bobs are both displaced by 12° .

Solution:

The movement of the pendula will not differ at all because the mass of the bob has no effect on the motion of a simple pendulum. The pendula are only affected by the period (which is related to the pendulum's length) and by the acceleration due to gravity.

Note:**PhET Explorations: Pendulum Lab**

Play with one or two pendulums and discover how the period of a simple pendulum depends on the length of the string, the mass of the pendulum bob, and the amplitude of the swing. It's easy to measure the period using the photogate timer. You can vary friction and the strength of gravity. Use the pendulum to find the value of g on planet X. Notice the anharmonic behavior at large amplitude.

https://phet.colorado.edu/sims/pendulum-lab/pendulum-lab_en.html

Section Summary

- A mass m suspended by a wire of length L is a simple pendulum and undergoes simple harmonic motion for amplitudes less than about 15° .

The period of a simple pendulum is

Equation:

$$T = 2\pi\sqrt{\frac{L}{g}},$$

where L is the length of the string and g is the acceleration due to gravity.

Conceptual Questions

Exercise:

Problem:

Pendulum clocks are made to run at the correct rate by adjusting the pendulum's length. Suppose you move from one city to another where the acceleration due to gravity is slightly greater, taking your pendulum clock with you, will you have to lengthen or shorten the pendulum to keep the correct time, other factors remaining constant? Explain your answer.

Problems & Exercises

As usual, the acceleration due to gravity in these problems is taken to be $g = 9.80 \text{ m/s}^2$, unless otherwise specified.

Exercise:

Problem:

What is the length of a pendulum that has a period of 0.500 s?

Solution:

6.21 cm

Exercise:

Problem:

Some people think a pendulum with a period of 1.00 s can be driven with “mental energy” or psycho kinetically, because its period is the same as an average heartbeat. True or not, what is the length of such a pendulum?

Exercise:

Problem: What is the period of a 1.00-m-long pendulum?

Solution:

2.01 s

Exercise:**Problem:**

How long does it take a child on a swing to complete one swing if her center of gravity is 4.00 m below the pivot?

Exercise:**Problem:**

The pendulum on a cuckoo clock is 5.00 cm long. What is its frequency?

Solution:

2.23 Hz

Exercise:**Problem:**

Two parakeets sit on a swing with their combined center of mass 10.0 cm below the pivot. At what frequency do they swing?

Exercise:

Problem:

(a) A pendulum that has a period of 3.00000 s and that is located where the acceleration due to gravity is 9.79 m/s^2 is moved to a location where the acceleration due to gravity is 9.82 m/s^2 . What is its new period? (b) Explain why so many digits are needed in the value for the period, based on the relation between the period and the acceleration due to gravity.

Solution:

(a) 2.99541 s

(b) Since the period is related to the square root of the acceleration of gravity, when the acceleration changes by 1% the period changes by $(0.01)^2 = 0.01\%$ so it is necessary to have at least 4 digits after the decimal to see the changes.

Exercise:**Problem:**

A pendulum with a period of 2.00000 s in one location ($g = 9.80 \text{ m/s}^2$) is moved to a new location where the period is now 1.99796 s. What is the acceleration due to gravity at its new location?

Exercise:**Problem:**

(a) What is the effect on the period of a pendulum if you double its length?

(b) What is the effect on the period of a pendulum if you decrease its length by 5.00%?

Solution:

(a) Period increases by a factor of 1.41 ($\sqrt{2}$)

(b) Period decreases to 97.5% of old period

Exercise:

Problem:

Find the ratio of the new/old periods of a pendulum if the pendulum were transported from Earth to the Moon, where the acceleration due to gravity is 1.63 m/s^2 .

Exercise:

Problem:

At what rate will a pendulum clock run on the Moon, where the acceleration due to gravity is 1.63 m/s^2 , if it keeps time accurately on Earth? That is, find the time (in hours) it takes the clock's hour hand to make one revolution on the Moon.

Solution:

Slow by a factor of 2.45

Exercise:

Problem:

Suppose the length of a clock's pendulum is changed by 1.000%, exactly at noon one day. What time will it read 24.00 hours later, assuming it the pendulum has kept perfect time before the change? Note that there are two answers, and perform the calculation to four-digit precision.

Exercise:

Problem:

If a pendulum-driven clock gains 5.00 s/day, what fractional change in pendulum length must be made for it to keep perfect time?

Solution:

length must increase by 0.0116%.

Glossary

simple pendulum

an object with a small mass suspended from a light wire or string

Energy and the Simple Harmonic Oscillator

- Determine the maximum speed of an oscillating system.

To study the energy of a simple harmonic oscillator, we first consider all the forms of energy it can have. We know from [Hooke's Law: Stress and Strain Revisited](#) that the energy stored in the deformation of a simple harmonic oscillator is a form of potential energy given by:

Equation:

$$PE_{\text{el}} = \frac{1}{2}kx^2.$$

Because a simple harmonic oscillator has no dissipative forces, the other important form of energy is kinetic energy KE. Conservation of energy for these two forms is:

Equation:

$$KE + PE_{\text{el}} = \text{constant}$$

or

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant}.$$

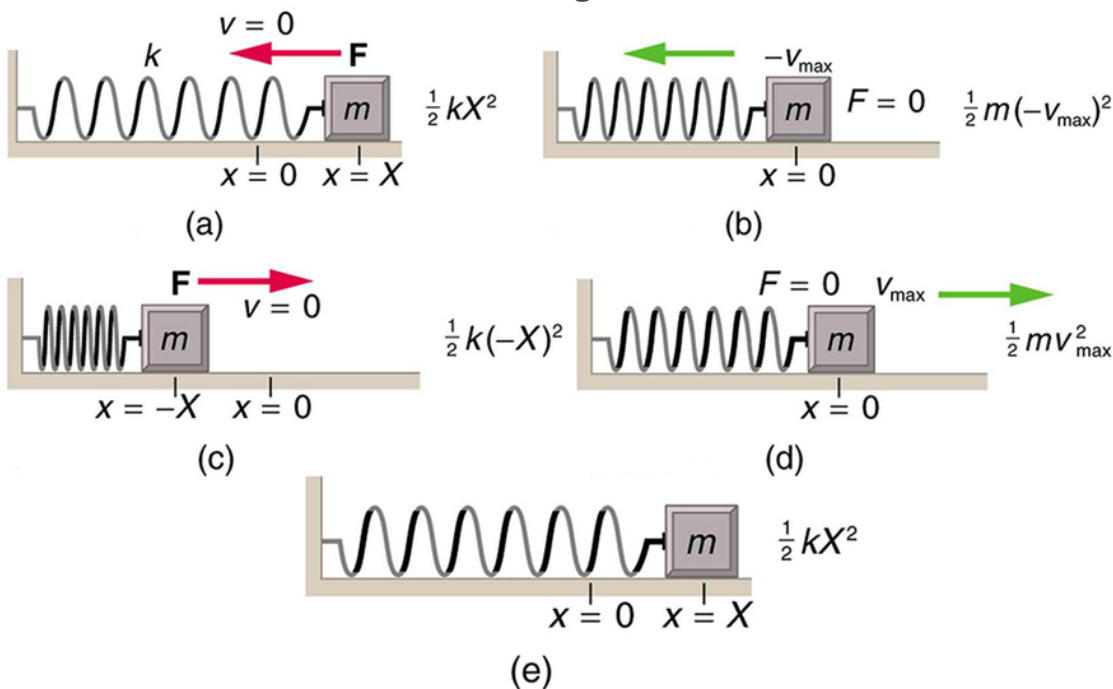
This statement of conservation of energy is valid for *all* simple harmonic oscillators, including ones where the gravitational force plays a role

Namely, for a simple pendulum we replace the velocity with $v = L\omega$, the spring constant with $k = mg/L$, and the displacement term with $x = L\theta$. Thus

Equation:

$$\frac{1}{2}mL^2\omega^2 + \frac{1}{2}mgL\theta^2 = \text{constant}.$$

In the case of undamped simple harmonic motion, the energy oscillates back and forth between kinetic and potential, going completely from one to the other as the system oscillates. So for the simple example of an object on a frictionless surface attached to a spring, as shown again in [\[link\]](#), the motion starts with all of the energy stored in the spring. As the object starts to move, the elastic potential energy is converted to kinetic energy, becoming entirely kinetic energy at the equilibrium position. It is then converted back into elastic potential energy by the spring, the velocity becomes zero when the kinetic energy is completely converted, and so on. This concept provides extra insight here and in later applications of simple harmonic motion, such as alternating current circuits.



The transformation of energy in simple harmonic motion is illustrated for an object attached to a spring on a frictionless surface.

The conservation of energy principle can be used to derive an expression for velocity v . If we start our simple harmonic motion with zero velocity and maximum displacement ($x = X$), then the total energy is

Equation:

$$\frac{1}{2}kX^2.$$

This total energy is constant and is shifted back and forth between kinetic energy and potential energy, at most times being shared by each. The conservation of energy for this system in equation form is thus:

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kX^2.$$

Solving this equation for v yields:

Equation:

$$v = \pm \sqrt{\frac{k}{m}(X^2 - x^2)}.$$

Manipulating this expression algebraically gives:

Equation:

$$v = \pm \sqrt{\frac{k}{m}}X\sqrt{1 - \frac{x^2}{X^2}}$$

and so

Equation:

$$v = \pm v_{\max} \sqrt{1 - \frac{x^2}{X^2}},$$

where

Equation:

$$v_{\max} = \sqrt{\frac{k}{m}} X.$$

From this expression, we see that the velocity is a maximum (v_{\max}) at $x = 0$, as stated earlier in $v(t) = -v_{\max} \sin \frac{2\pi t}{T}$. Notice that the maximum velocity depends on three factors. Maximum velocity is directly proportional to amplitude. As you might guess, the greater the maximum displacement the greater the maximum velocity. Maximum velocity is also greater for stiffer systems, because they exert greater force for the same displacement. This observation is seen in the expression for v_{\max} ; it is proportional to the square root of the force constant k . Finally, the maximum velocity is smaller for objects that have larger masses, because the maximum velocity is inversely proportional to the square root of m . For a given force, objects that have large masses accelerate more slowly.

A similar calculation for the simple pendulum produces a similar result, namely:

Equation:

$$\omega_{\max} = \sqrt{\frac{g}{L}} \theta_{\max}.$$

Example:

Determine the Maximum Speed of an Oscillating System: A Bumpy Road

Suppose that a car is 900 kg and has a suspension system that has a force constant $k = 6.53 \times 10^4$ N/m. The car hits a bump and bounces with an amplitude of 0.100 m. What is its maximum vertical velocity if you assume no damping occurs?

Strategy

We can use the expression for v_{\max} given in $v_{\max} = \sqrt{\frac{k}{m}} X$ to determine the maximum vertical velocity. The variables m and k are given in the

problem statement, and the maximum displacement X is 0.100 m.

Solution

1. Identify known.

2. Substitute known values into $v_{\max} = \sqrt{\frac{k}{m}} X$:

Equation:

$$v_{\max} = \sqrt{\frac{6.53 \times 10^4 \text{ N/m}}{900 \text{ kg}}} (0.100 \text{ m}).$$

3. Calculate to find $v_{\max} = 0.852 \text{ m/s}$.

Discussion

This answer seems reasonable for a bouncing car. There are other ways to use conservation of energy to find v_{\max} . We could use it directly, as was done in the example featured in [Hooke's Law: Stress and Strain Revisited](#).

The small vertical displacement y of an oscillating simple pendulum, starting from its equilibrium position, is given as

Equation:

$$y(t) = a \sin \omega t,$$

where a is the amplitude, ω is the angular velocity and t is the time taken. Substituting $\omega = \frac{2\pi}{T}$, we have

Equation:

$$y(t) = a \sin \left(\frac{2\pi t}{T} \right).$$

Thus, the displacement of pendulum is a function of time as shown above. Also the velocity of the pendulum is given by

Equation:

$$v(t) = \frac{2a\pi}{T} \cos \left(\frac{2\pi t}{T} \right),$$

so the motion of the pendulum is a function of time.

Exercise:

Check Your Understanding

Problem:

Why does it hurt more if your hand is snapped with a ruler than with a loose spring, even if the displacement of each system is equal?

Solution:

The ruler is a stiffer system, which carries greater force for the same amount of displacement. The ruler snaps your hand with greater force, which hurts more.

Exercise:

Check Your Understanding

Problem:

You are observing a simple harmonic oscillator. Identify one way you could decrease the maximum velocity of the system.

Solution:

You could increase the mass of the object that is oscillating.

Section Summary

- Energy in the simple harmonic oscillator is shared between elastic potential energy and kinetic energy, with the total being constant:

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant}.$$

- Maximum velocity depends on three factors: it is directly proportional to amplitude, it is greater for stiffer systems, and it is smaller for objects that have larger masses:

Equation:

$$v_{\max} = \sqrt{\frac{k}{m}} X.$$

Conceptual Questions

Exercise:

Problem:

Explain in terms of energy how dissipative forces such as friction reduce the amplitude of a harmonic oscillator. Also explain how a driving mechanism can compensate. (A pendulum clock is such a system.)

Problems & Exercises

Exercise:

Problem:

The length of nylon rope from which a mountain climber is suspended has a force constant of $1.40 \times 10^4 \text{ N/m}$.

- (a) What is the frequency at which he bounces, given his mass plus and the mass of his equipment are 90.0 kg?
- (b) How much would this rope stretch to break the climber's fall if he free-falls 2.00 m before the rope runs out of slack? Hint: Use conservation of energy.
- (c) Repeat both parts of this problem in the situation where twice this length of nylon rope is used.

Solution:

(a) 1.99 Hz

(b) 50.2 cm

(c) 1.41 Hz, 0.710 m

Exercise:**Problem: Engineering Application**

Near the top of the Citigroup Center building in New York City, there is an object with mass of 4.00×10^5 kg on springs that have adjustable force constants. Its function is to dampen wind-driven oscillations of the building by oscillating at the same frequency as the building is being driven—the driving force is transferred to the object, which oscillates instead of the entire building. (a) What effective force constant should the springs have to make the object oscillate with a period of 2.00 s? (b) What energy is stored in the springs for a 2.00-m displacement from equilibrium?

Solution:

(a) 3.95×10^6 N/m

(b) 7.90×10^6 J

Forced Oscillations and Resonance

- Observe resonance of a paddle ball on a string.
- Observe amplitude of a damped harmonic oscillator.

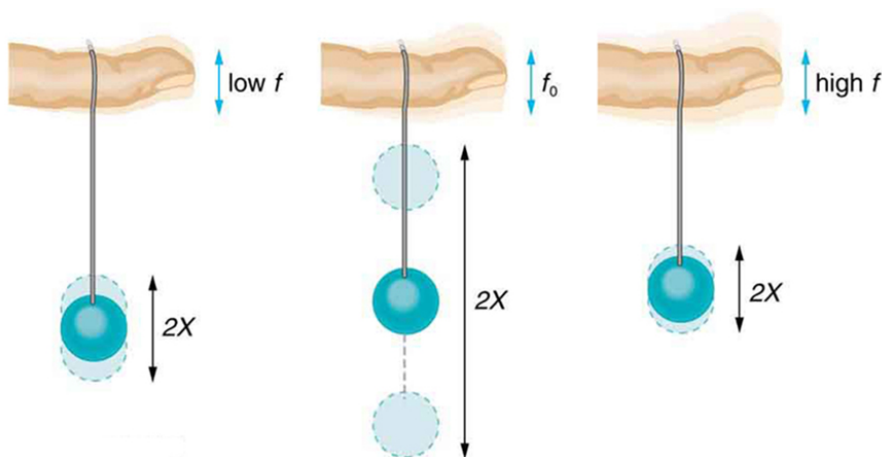


You can cause the strings in a piano to vibrate simply by producing sound waves from your voice. (credit: Matt Billings, Flickr)

Sit in front of a piano sometime and sing a loud brief note at it with the dampers off its strings. It will sing the same note back at you—the strings, having the same frequencies as your voice, are resonating in response to the forces from the sound waves that you sent to them. Your voice and a piano's strings is a good example of the fact that objects—in this case, piano strings—can be forced to oscillate but oscillate best at their natural frequency. In this section, we shall briefly explore applying a *periodic driving force* acting on a simple harmonic oscillator. The driving force puts energy into the system at a certain frequency, not necessarily the same as the natural frequency of the system. The **natural frequency** is the frequency at which a system would oscillate if there were no driving and no damping force.

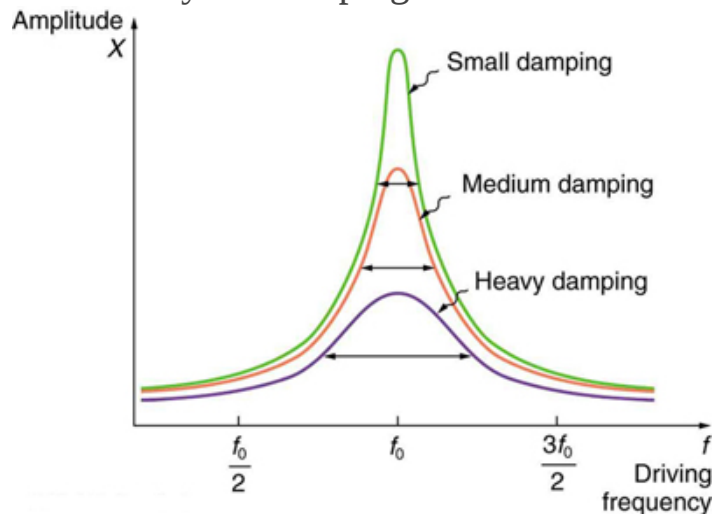
Most of us have played with toys involving an object supported on an elastic band, something like the paddle ball suspended from a finger in [\[link\]](#). Imagine the finger in the figure is your finger. At first you hold your

finger steady, and the ball bounces up and down with a small amount of damping. If you move your finger up and down slowly, the ball will follow along without bouncing much on its own. As you increase the frequency at which you move your finger up and down, the ball will respond by oscillating with increasing amplitude. When you drive the ball at its natural frequency, the ball's oscillations increase in amplitude with each oscillation for as long as you drive it. The phenomenon of driving a system with a frequency equal to its natural frequency is called **resonance**. A system being driven at its natural frequency is said to **resonate**. As the driving frequency gets progressively higher than the resonant or natural frequency, the amplitude of the oscillations becomes smaller, until the oscillations nearly disappear and your finger simply moves up and down with little effect on the ball.



The paddle ball on its rubber band moves in response to the finger supporting it. If the finger moves with the natural frequency f_0 of the ball on the rubber band, then a resonance is achieved, and the amplitude of the ball's oscillations increases dramatically. At higher and lower driving frequencies, energy is transferred to the ball less efficiently, and it responds with lower-amplitude oscillations.

[\[link\]](#) shows a graph of the amplitude of a damped harmonic oscillator as a function of the frequency of the periodic force driving it. There are three curves on the graph, each representing a different amount of damping. All three curves peak at the point where the frequency of the driving force equals the natural frequency of the harmonic oscillator. The highest peak, or greatest response, is for the least amount of damping, because less energy is removed by the damping force.



Amplitude of a harmonic oscillator as a function of the frequency of the driving force. The curves represent the same oscillator with the same natural frequency but with different amounts of damping. Resonance occurs when the driving frequency equals the natural frequency, and the greatest response is for the least amount of damping. The narrowest response is also for the least damping.

It is interesting that the widths of the resonance curves shown in [\[link\]](#) depend on damping: the less the damping, the narrower the resonance. The message is that if you want a driven oscillator to resonate at a very specific frequency, you need as little damping as possible. Little damping is the case

for piano strings and many other musical instruments. Conversely, if you want small-amplitude oscillations, such as in a car's suspension system, then you want heavy damping. Heavy damping reduces the amplitude, but the tradeoff is that the system responds at more frequencies.

These features of driven harmonic oscillators apply to a huge variety of systems. When you tune a radio, for example, you are adjusting its resonant frequency so that it only oscillates to the desired station's broadcast (driving) frequency. The more selective the radio is in discriminating between stations, the smaller its damping. Magnetic resonance imaging (MRI) is a widely used medical diagnostic tool in which atomic nuclei (mostly hydrogen nuclei) are made to resonate by incoming radio waves (on the order of 100 MHz). A child on a swing is driven by a parent at the swing's natural frequency to achieve maximum amplitude. In all of these cases, the efficiency of energy transfer from the driving force into the oscillator is best at resonance. Speed bumps and gravel roads prove that even a car's suspension system is not immune to resonance. In spite of finely engineered shock absorbers, which ordinarily convert mechanical energy to thermal energy almost as fast as it comes in, speed bumps still cause a large-amplitude oscillation. On gravel roads that are corrugated, you may have noticed that if you travel at the "wrong" speed, the bumps are very noticeable whereas at other speeds you may hardly feel the bumps at all. [\[link\]](#) shows a photograph of a famous example (the Tacoma Narrows Bridge) of the destructive effects of a driven harmonic oscillation. The Millennium Bridge in London was closed for a short period of time for the same reason while inspections were carried out.

In our bodies, the chest cavity is a clear example of a system at resonance. The diaphragm and chest wall drive the oscillations of the chest cavity which result in the lungs inflating and deflating. The system is critically damped and the muscular diaphragm oscillates at the resonant value for the system, making it highly efficient.



In 1940, the Tacoma Narrows Bridge in Washington state collapsed. Heavy cross winds drove the bridge into oscillations at its resonant frequency. Damping decreased when support cables broke loose and started to slip over the towers, allowing increasingly greater amplitudes until the structure failed (credit: PRI's *Studio 360*, via Flickr)

Exercise:
Check Your Understanding

Problem:

A famous magic trick involves a performer singing a note toward a crystal glass until the glass shatters. Explain why the trick works in terms of resonance and natural frequency.

Solution:

The performer must be singing a note that corresponds to the natural frequency of the glass. As the sound wave is directed at the glass, the glass responds by resonating at the same frequency as the sound wave.

With enough energy introduced into the system, the glass begins to vibrate and eventually shatters.

Section Summary

- A system's natural frequency is the frequency at which the system will oscillate if not affected by driving or damping forces.
- A periodic force driving a harmonic oscillator at its natural frequency produces resonance. The system is said to resonate.
- The less damping a system has, the higher the amplitude of the forced oscillations near resonance. The more damping a system has, the broader response it has to varying driving frequencies.

Conceptual Questions

Exercise:

Problem:

Why are soldiers in general ordered to “route step” (walk out of step) across a bridge?

Problems & Exercises

Exercise:

Problem:

How much energy must the shock absorbers of a 1200-kg car dissipate in order to damp a bounce that initially has a velocity of 0.800 m/s at the equilibrium position? Assume the car returns to its original vertical position.

Solution:

384 J

Exercise:**Problem:**

If a car has a suspension system with a force constant of $5.00 \times 10^4 \text{ N/m}$, how much energy must the car's shocks remove to dampen an oscillation starting with a maximum displacement of 0.0750 m?

Exercise:**Problem:**

(a) How much will a spring that has a force constant of 40.0 N/m be stretched by an object with a mass of 0.500 kg when hung motionless from the spring? (b) Calculate the decrease in gravitational potential energy of the 0.500-kg object when it descends this distance. (c) Part of this gravitational energy goes into the spring. Calculate the energy stored in the spring by this stretch, and compare it with the gravitational potential energy. Explain where the rest of the energy might go.

Solution:

(a). 0.123 m

(b). -0.600 J

(c). 0.300 J. The rest of the energy may go into heat caused by friction and other damping forces.

Exercise:

Problem:

Suppose you have a 0.750-kg object on a horizontal surface connected to a spring that has a force constant of 150 N/m. There is simple friction between the object and surface with a static coefficient of friction $\mu_s = 0.100$. (a) How far can the spring be stretched without moving the mass? (b) If the object is set into oscillation with an amplitude twice the distance found in part (a), and the kinetic coefficient of friction is $\mu_k = 0.0850$, what total distance does it travel before stopping? Assume it starts at the maximum amplitude.

Exercise:**Problem:**

Engineering Application: A suspension bridge oscillates with an effective force constant of 1.00×10^8 N/m. (a) How much energy is needed to make it oscillate with an amplitude of 0.100 m? (b) If soldiers march across the bridge with a cadence equal to the bridge's natural frequency and impart 1.00×10^4 J of energy each second, how long does it take for the bridge's oscillations to go from 0.100 m to 0.500 m amplitude?

Solution:

(a) 5.00×10^5 J

(b) 1.20×10^3 s

Glossary

natural frequency

the frequency at which a system would oscillate if there were no driving and no damping forces

resonance

the phenomenon of driving a system with a frequency equal to the system's natural frequency

resonate

a system being driven at its natural frequency

Waves

- State the characteristics of a wave.
- Calculate the velocity of wave propagation.



Waves in the ocean behave similarly to all other types of waves. (credit: Steve Jurveston, Flickr)

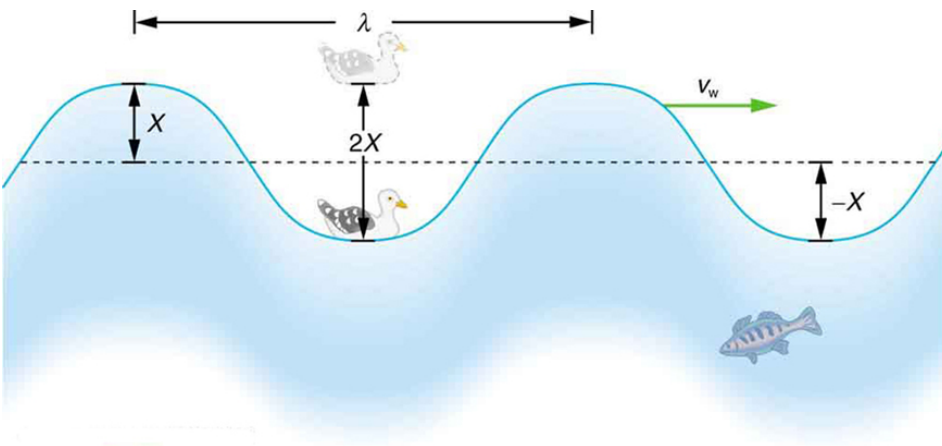
What do we mean when we say something is a wave? The most intuitive and easiest wave to imagine is the familiar water wave. More precisely, a **wave** is a disturbance that propagates, or moves from the place it was created. For water waves, the disturbance is in the surface of the water, perhaps created by a rock thrown into a pond or by a swimmer splashing the surface repeatedly. For sound waves, the disturbance is a change in air pressure, perhaps created by the oscillating cone inside a speaker. For earthquakes, there are several types of disturbances, including disturbance of Earth's surface and pressure disturbances under the surface. Even radio waves are most easily understood using an analogy with water waves. Visualizing water waves is useful because there is more to it than just a mental image. Water waves exhibit characteristics common to all waves, such as amplitude, period, frequency and energy. All wave characteristics can be described by a small set of underlying principles.

A wave is a disturbance that propagates, or moves from the place it was created. The simplest waves repeat themselves for several cycles and are associated with simple harmonic motion. Let us start by considering the simplified water wave in [\[link\]](#). The wave is an up and down disturbance of the water surface. It causes a sea gull to move up and down in simple harmonic motion as the wave crests and troughs (peaks and valleys) pass under the bird. The time for one complete up and down motion is the wave's period T . The wave's frequency is $f = 1/T$, as usual. The wave itself moves to the right in the figure. This movement of the wave is actually the disturbance moving to the right, not the water itself (or the bird would move to the right). We define **wave velocity** v_w to be the speed at which the disturbance moves. Wave velocity is sometimes also called the *propagation velocity* or *propagation speed*, because the disturbance propagates from one location to another.

Note:

Misconception Alert

Many people think that water waves push water from one direction to another. In fact, the particles of water tend to stay in one location, save for moving up and down due to the energy in the wave. The energy moves forward through the water, but the water stays in one place. If you feel yourself pushed in an ocean, what you feel is the energy of the wave, not a rush of water.



An idealized ocean wave passes under a sea gull that bobs up and down in simple harmonic motion. The wave has a wavelength λ , which is the distance between adjacent identical parts of the wave. The up and down disturbance of the surface propagates parallel to the surface at a speed v_w .

The water wave in the figure also has a length associated with it, called its **wavelength** λ , the distance between adjacent identical parts of a wave. (λ is the distance parallel to the direction of propagation.) The speed of propagation v_w is the distance the wave travels in a given time, which is one wavelength in the time of one period. In equation form, that is

Equation:

$$v_w = \frac{\lambda}{T}$$

or

Equation:

$$v_w = f\lambda.$$

This fundamental relationship holds for all types of waves. For water waves, v_w is the speed of a surface wave; for sound, v_w is the speed of sound; and for visible light, v_w is the speed of light, for example.

Note:

Take-Home Experiment: Waves in a Bowl

Fill a large bowl or basin with water and wait for the water to settle so there are no ripples. Gently drop a cork into the middle of the bowl. Estimate the wavelength and period of oscillation of the water wave that propagates away from the cork. Remove the cork from the bowl and wait

for the water to settle again. Gently drop the cork at a height that is different from the first drop. Does the wavelength depend upon how high above the water the cork is dropped?

Example:

Calculate the Velocity of Wave Propagation: Gull in the Ocean

Calculate the wave velocity of the ocean wave in [\[link\]](#) if the distance between wave crests is 10.0 m and the time for a sea gull to bob up and down is 5.00 s.

Strategy

We are asked to find v_w . The given information tells us that $\lambda = 10.0$ m and $T = 5.00$ s. Therefore, we can use $v_w = \frac{\lambda}{T}$ to find the wave velocity.

Solution

1. Enter the known values into $v_w = \frac{\lambda}{T}$:

Equation:

$$v_w = \frac{10.0 \text{ m}}{5.00 \text{ s}}.$$

2. Solve for v_w to find $v_w = 2.00$ m/s.

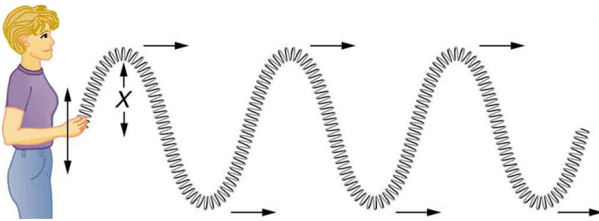
Discussion

This slow speed seems reasonable for an ocean wave. Note that the wave moves to the right in the figure at this speed, not the varying speed at which the sea gull moves up and down.

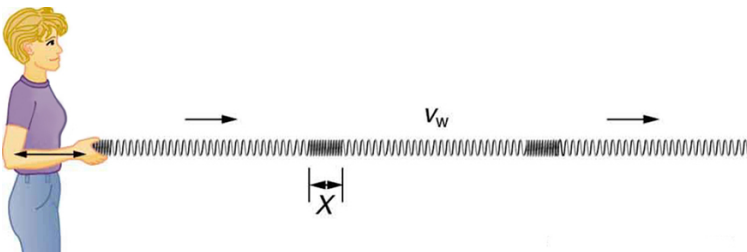
Transverse and Longitudinal Waves

A simple wave consists of a periodic disturbance that propagates from one place to another. The wave in [\[link\]](#) propagates in the horizontal direction while the surface is disturbed in the vertical direction. Such a wave is called a **transverse wave** or shear wave; in such a wave, the disturbance is perpendicular to the direction of propagation. In contrast, in a **longitudinal**

wave or compressional wave, the disturbance is parallel to the direction of propagation. [\[link\]](#) shows an example of a longitudinal wave. The size of the disturbance is its amplitude X and is completely independent of the speed of propagation v_w .



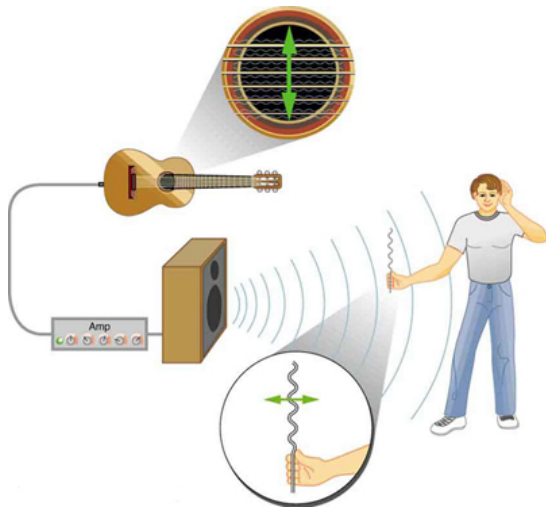
In this example of a transverse wave, the wave propagates horizontally, and the disturbance in the cord is in the vertical direction.



In this example of a longitudinal wave, the wave propagates horizontally, and the disturbance in the cord is also in the horizontal direction.

Waves may be transverse, longitudinal, or *a combination of the two*. (Water waves are actually a combination of transverse and longitudinal. The simplified water wave illustrated in [\[link\]](#) shows no longitudinal motion of the bird.) The waves on the strings of musical instruments are transverse—so are electromagnetic waves, such as visible light.

Sound waves in air and water are longitudinal. Their disturbances are periodic variations in pressure that are transmitted in fluids. Fluids do not have appreciable shear strength, and thus the sound waves in them must be longitudinal or compressional. Sound in solids can be both longitudinal and transverse.



The wave on a guitar string is transverse. The sound wave rattles a sheet of paper in a direction that shows the sound wave is longitudinal.

Earthquake waves under Earth's surface also have both longitudinal and transverse components (called compressional or P-waves and shear or S-waves, respectively). These components have important individual characteristics—they propagate at different speeds, for example.

Earthquakes also have surface waves that are similar to surface waves on water.

Exercise:

Check Your Understanding

Problem:

Why is it important to differentiate between longitudinal and transverse waves?

Solution:

In the different types of waves, energy can propagate in a different direction relative to the motion of the wave. This is important to understand how different types of waves affect the materials around them.

Note:

PhET Explorations: Wave on a String

Watch a string vibrate in slow motion. Wiggle the end of the string and make waves, or adjust the frequency and amplitude of an oscillator. Adjust the damping and tension. The end can be fixed, loose, or open.

https://phet.colorado.edu/sims/html/wave-on-a-string/latest/wave-on-a-string_en.html

Section Summary

- A wave is a disturbance that moves from the point of creation with a wave velocity v_w .
- A wave has a wavelength λ , which is the distance between adjacent identical parts of the wave.
- Wave velocity and wavelength are related to the wave's frequency and period by $v_w = \frac{\lambda}{T}$ or $v_w = f\lambda$.

- A transverse wave has a disturbance perpendicular to its direction of propagation, whereas a longitudinal wave has a disturbance parallel to its direction of propagation.

Conceptual Questions

Exercise:

Problem:

Give one example of a transverse wave and another of a longitudinal wave, being careful to note the relative directions of the disturbance and wave propagation in each.

Exercise:

Problem:

What is the difference between propagation speed and the frequency of a wave? Does one or both affect wavelength? If so, how?

Problems & Exercises

Exercise:

Problem:

Storms in the South Pacific can create waves that travel all the way to the California coast, which are 12,000 km away. How long does it take them if they travel at 15.0 m/s?

Solution:

Equation:

$$t = 9.26 \text{ d}$$

Exercise:

Problem:

Waves on a swimming pool propagate at 0.750 m/s. You splash the water at one end of the pool and observe the wave go to the opposite end, reflect, and return in 30.0 s. How far away is the other end of the pool?

Exercise:**Problem:**

Wind gusts create ripples on the ocean that have a wavelength of 5.00 cm and propagate at 2.00 m/s. What is their frequency?

Solution:**Equation:**

$$f = 40.0 \text{ Hz}$$

Exercise:**Problem:**

How many times a minute does a boat bob up and down on ocean waves that have a wavelength of 40.0 m and a propagation speed of 5.00 m/s?

Exercise:**Problem:**

Scouts at a camp shake the rope bridge they have just crossed and observe the wave crests to be 8.00 m apart. If they shake it the bridge twice per second, what is the propagation speed of the waves?

Solution:**Equation:**

$$v_w = 16.0 \text{ m/s}$$

Exercise:**Problem:**

What is the wavelength of the waves you create in a swimming pool if you splash your hand at a rate of 2.00 Hz and the waves propagate at 0.800 m/s?

Exercise:**Problem:**

What is the wavelength of an earthquake that shakes you with a frequency of 10.0 Hz and gets to another city 84.0 km away in 12.0 s?

Solution:**Equation:**

$$\lambda = 700 \text{ m}$$

Exercise:**Problem:**

Radio waves transmitted through space at $3.00 \times 10^8 \text{ m/s}$ by the Voyager spacecraft have a wavelength of 0.120 m. What is their frequency?

Exercise:**Problem:**

Your ear is capable of differentiating sounds that arrive at the ear just 1.00 ms apart. What is the minimum distance between two speakers that produce sounds that arrive at noticeably different times on a day when the speed of sound is 340 m/s?

Solution:**Equation:**

$$d = 34.0 \text{ cm}$$

Exercise:

Problem:

(a) Seismographs measure the arrival times of earthquakes with a precision of 0.100 s. To get the distance to the epicenter of the quake, they compare the arrival times of S- and P-waves, which travel at different speeds. ([link](#)) If S- and P-waves travel at 4.00 and 7.20 km/s, respectively, in the region considered, how precisely can the distance to the source of the earthquake be determined? (b) Seismic waves from underground detonations of nuclear bombs can be used to locate the test site and detect violations of test bans. Discuss whether your answer to (a) implies a serious limit to such detection. (Note also that the uncertainty is greater if there is an uncertainty in the propagation speeds of the S- and P-waves.)



A seismograph as described in above problem.(credit: Oleg Alexandrov)

Glossary

longitudinal wave

a wave in which the disturbance is parallel to the direction of propagation

transverse wave

a wave in which the disturbance is perpendicular to the direction of propagation

wave velocity

the speed at which the disturbance moves. Also called the propagation velocity or propagation speed

wavelength

the distance between adjacent identical parts of a wave

Superposition and Interference

- Explain standing waves.
- Describe the mathematical representation of overtones and beat frequency.



These waves result from the superposition of several waves from different sources, producing a complex pattern.
(credit: waterborough, Wikimedia Commons)

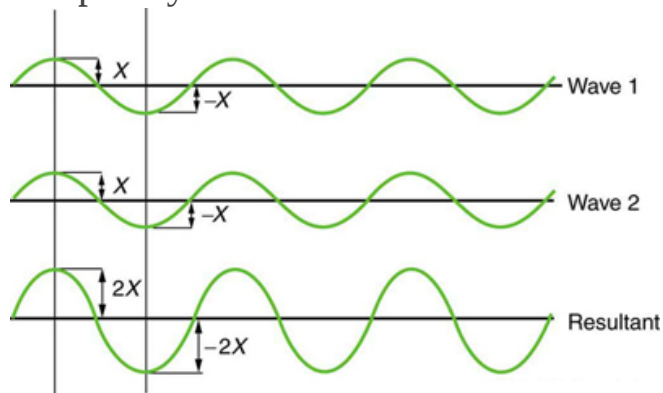
Most waves do not look very simple. They look more like the waves in [\[link\]](#) than like the simple water wave considered in [Waves](#). (Simple waves may be created by a simple harmonic oscillation, and thus have a sinusoidal shape). Complex waves are more interesting, even beautiful, but they look formidable. Most waves appear complex because they result from several simple waves adding together. Luckily, the rules for adding waves are quite simple.

When two or more waves arrive at the same point, they superimpose themselves on one another. More specifically, the disturbances of waves are superimposed when they come together—a phenomenon called **superposition**. Each disturbance corresponds to a force, and forces add. If the disturbances are along the same line, then the resulting wave is a simple

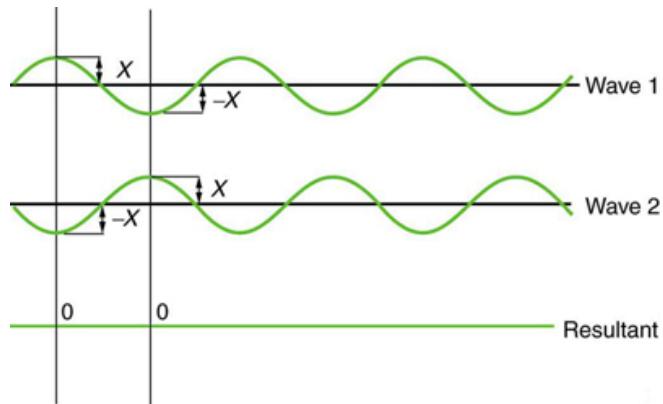
addition of the disturbances of the individual waves—that is, their amplitudes add. [\[link\]](#) and [\[link\]](#) illustrate superposition in two special cases, both of which produce simple results.

[\[link\]](#) shows two identical waves that arrive at the same point exactly in phase. The crests of the two waves are precisely aligned, as are the troughs. This superposition produces pure **constructive interference**. Because the disturbances add, pure constructive interference produces a wave that has twice the amplitude of the individual waves, but has the same wavelength.

[\[link\]](#) shows two identical waves that arrive exactly out of phase—that is, precisely aligned crest to trough—producing pure **destructive interference**. Because the disturbances are in the opposite direction for this superposition, the resulting amplitude is zero for pure destructive interference—the waves completely cancel.



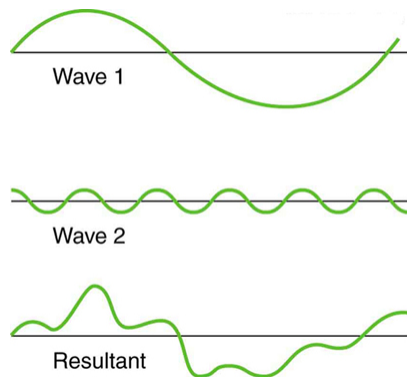
Pure constructive interference of two identical waves produces one with twice the amplitude, but the same wavelength.



Pure destructive interference of two identical waves produces zero amplitude, or complete cancellation.

While pure constructive and pure destructive interference do occur, they require precisely aligned identical waves. The superposition of most waves produces a combination of constructive and destructive interference and can vary from place to place and time to time. Sound from a stereo, for example, can be loud in one spot and quiet in another. Varying loudness means the sound waves add partially constructively and partially destructively at different locations. A stereo has at least two speakers creating sound waves, and waves can reflect from walls. All these waves superimpose. An example of sounds that vary over time from constructive to destructive is found in the combined whine of airplane jets heard by a stationary passenger. The combined sound can fluctuate up and down in volume as the sound from the two engines varies in time from constructive to destructive. These examples are of waves that are similar.

An example of the superposition of two dissimilar waves is shown in [\[link\]](#). Here again, the disturbances add and subtract, producing a more complicated looking wave.

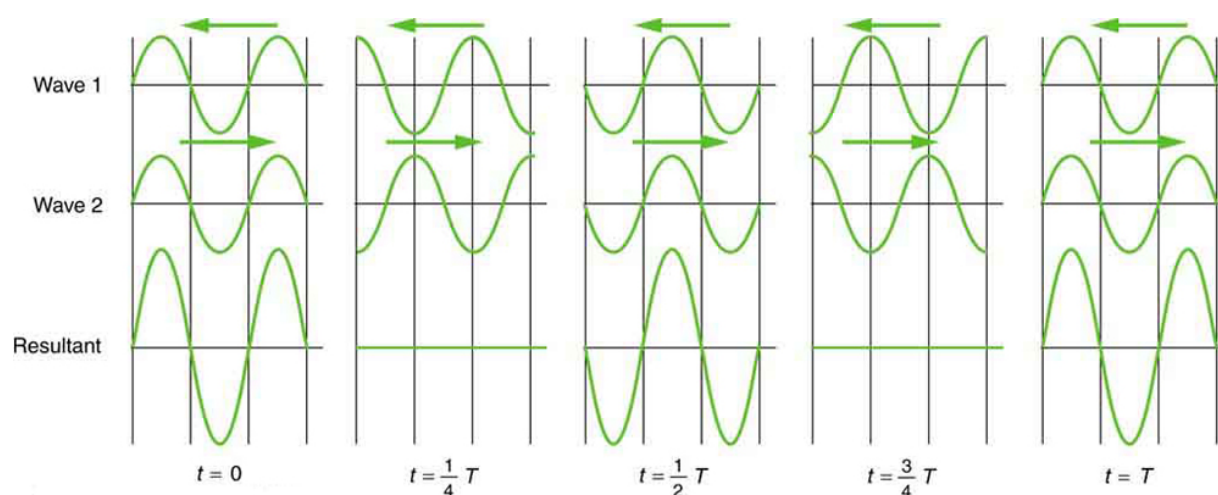


Superposition of
non-identical
waves exhibits both
constructive and
destructive
interference.

Standing Waves

Sometimes waves do not seem to move; rather, they just vibrate in place. Unmoving waves can be seen on the surface of a glass of milk in a refrigerator, for example. Vibrations from the refrigerator motor create waves on the milk that oscillate up and down but do not seem to move across the surface. These waves are formed by the superposition of two or more moving waves, such as illustrated in [\[link\]](#) for two identical waves moving in opposite directions. The waves move through each other with their disturbances adding as they go by. If the two waves have the same amplitude and wavelength, then they alternate between constructive and destructive interference. The resultant looks like a wave standing in place and, thus, is called a **standing wave**. Waves on the glass of milk are one example of standing waves. There are other standing waves, such as on guitar strings and in organ pipes. With the glass of milk, the two waves that produce standing waves may come from reflections from the side of the glass.

A closer look at earthquakes provides evidence for conditions appropriate for resonance, standing waves, and constructive and destructive interference. A building may be vibrated for several seconds with a driving frequency matching that of the natural frequency of vibration of the building—producing a resonance resulting in one building collapsing while neighboring buildings do not. Often buildings of a certain height are devastated while other taller buildings remain intact. The building height matches the condition for setting up a standing wave for that particular height. As the earthquake waves travel along the surface of Earth and reflect off denser rocks, constructive interference occurs at certain points. Often areas closer to the epicenter are not damaged while areas farther away are damaged.

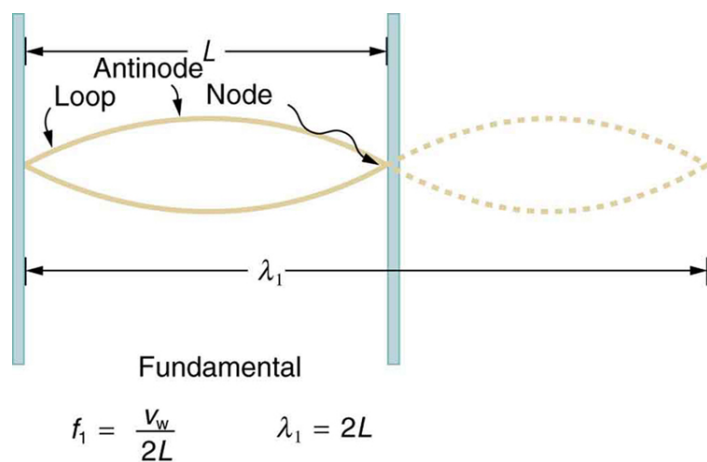


Standing wave created by the superposition of two identical waves moving in opposite directions. The oscillations are at fixed locations in space and result from alternately constructive and destructive interference.

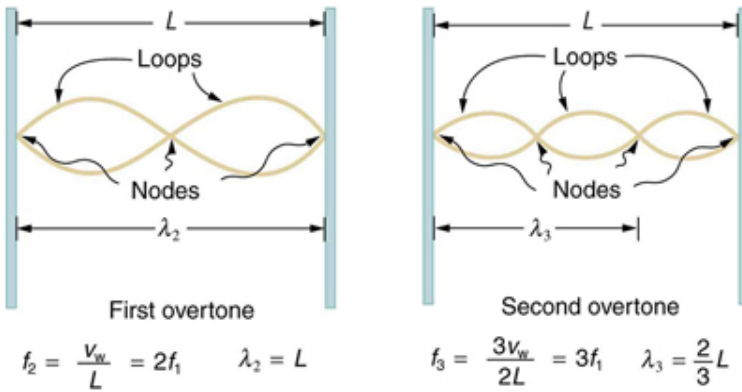
Standing waves are also found on the strings of musical instruments and are due to reflections of waves from the ends of the string. [\[link\]](#) and [\[link\]](#) show three standing waves that can be created on a string that is fixed at both ends. **Nodes** are the points where the string does not move; more

generally, nodes are where the wave disturbance is zero in a standing wave. The fixed ends of strings must be nodes, too, because the string cannot move there. The word **antinode** is used to denote the location of maximum amplitude in standing waves. Standing waves on strings have a frequency that is related to the propagation speed v_w of the disturbance on the string. The wavelength λ is determined by the distance between the points where the string is fixed in place.

The lowest frequency, called the **fundamental frequency**, is thus for the longest wavelength, which is seen to be $\lambda_1 = 2L$. Therefore, the fundamental frequency is $f_1 = v_w/\lambda_1 = v_w/2L$. In this case, the **overtones** or harmonics are multiples of the fundamental frequency. As seen in [\[link\]](#), the first harmonic can easily be calculated since $\lambda_2 = L$. Thus, $f_2 = v_w/\lambda_2 = v_w/2L = 2f_1$. Similarly, $f_3 = 3f_1$, and so on. All of these frequencies can be changed by adjusting the tension in the string. The greater the tension, the greater v_w is and the higher the frequencies. This observation is familiar to anyone who has ever observed a string instrument being tuned. We will see in later chapters that standing waves are crucial to many resonance phenomena, such as in sounding boxes on string instruments.



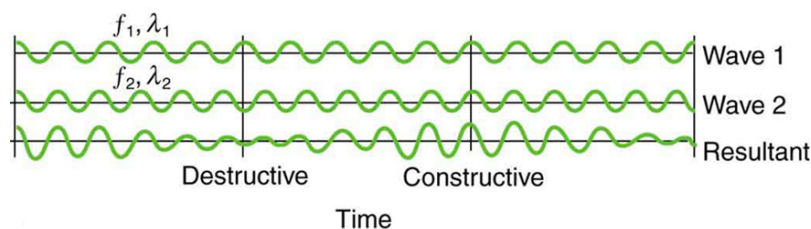
The figure shows a string oscillating at its fundamental frequency.



First and second harmonic frequencies are shown.

Beats

Striking two adjacent keys on a piano produces a warbling combination usually considered to be unpleasant. The superposition of two waves of similar but not identical frequencies is the culprit. Another example is often noticeable in jet aircraft, particularly the two-engine variety, while taxiing. The combined sound of the engines goes up and down in loudness. This varying loudness happens because the sound waves have similar but not identical frequencies. The discordant warbling of the piano and the fluctuating loudness of the jet engine noise are both due to alternately constructive and destructive interference as the two waves go in and out of phase. [\[link\]](#) illustrates this graphically.



Beats are produced by the superposition of two waves of slightly different frequencies but identical amplitudes. The waves alternate in time between constructive interference and destructive interference, giving the resulting wave a time-varying amplitude.

The wave resulting from the superposition of two similar-frequency waves has a frequency that is the average of the two. This wave fluctuates in amplitude, or *beats*, with a frequency called the **beat frequency**. We can determine the beat frequency by adding two waves together mathematically. Note that a wave can be represented at one point in space as

Equation:

$$x = X \cos\left(\frac{2\pi t}{T}\right) = X \cos(2\pi ft),$$

where $f = 1/T$ is the frequency of the wave. Adding two waves that have different frequencies but identical amplitudes produces a resultant

Equation:

$$x = x_1 + x_2.$$

More specifically,

Equation:

$$x = X \cos(2\pi f_1 t) + X \cos(2\pi f_2 t).$$

Using a trigonometric identity, it can be shown that

Equation:

$$x = 2X \cos(\pi f_B t) \cos(2\pi f_{\text{ave}} t),$$

where

Equation:

$$f_B = |f_1 - f_2|$$

is the beat frequency, and f_{ave} is the average of f_1 and f_2 . These results mean that the resultant wave has twice the amplitude and the average frequency of the two superimposed waves, but it also fluctuates in overall amplitude at the beat frequency f_B . The first cosine term in the expression effectively causes the amplitude to go up and down. The second cosine term is the wave with frequency f_{ave} . This result is valid for all types of waves. However, if it is a sound wave, providing the two frequencies are similar, then what we hear is an average frequency that gets louder and softer (or warbles) at the beat frequency.

Note:

Making Career Connections

Piano tuners use beats routinely in their work. When comparing a note with a tuning fork, they listen for beats and adjust the string until the beats go away (to zero frequency). For example, if the tuning fork has a 256 Hz frequency and two beats per second are heard, then the other frequency is either 254 or 258 Hz. Most keys hit multiple strings, and these strings are actually adjusted until they have nearly the same frequency and give a slow beat for richness. Twelve-string guitars and mandolins are also tuned using beats.

While beats may sometimes be annoying in audible sounds, we will find that beats have many applications. Observing beats is a very useful way to compare similar frequencies. There are applications of beats as apparently disparate as in ultrasonic imaging and radar speed traps.

Exercise:

Check Your Understanding

Problem:

Imagine you are holding one end of a jump rope, and your friend holds the other. If your friend holds her end still, you can move your end up and down, creating a transverse wave. If your friend then begins to move her end up and down, generating a wave in the opposite direction, what resultant wave forms would you expect to see in the jump rope?

Solution:

The rope would alternate between having waves with amplitudes two times the original amplitude and reaching equilibrium with no amplitude at all. The wavelengths will result in both constructive and destructive interference

Exercise:

Check Your Understanding

Problem: Define nodes and antinodes.

Solution:

Nodes are areas of wave interference where there is no motion. Antinodes are areas of wave interference where the motion is at its maximum point.

Exercise:

Check Your Understanding

Problem:

You hook up a stereo system. When you test the system, you notice that in one corner of the room, the sounds seem dull. In another area, the sounds seem excessively loud. Describe how the sound moving about the room could result in these effects.

Solution:

With multiple speakers putting out sounds into the room, and these sounds bouncing off walls, there is bound to be some wave interference. In the dull areas, the interference is probably mostly destructive. In the louder areas, the interference is probably mostly constructive.

Note:

PhET Explorations: Wave Interference

Make waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern.

[Wave](#)
[Interferenc](#)
[e](#)

Section Summary

- Superposition is the combination of two waves at the same location.
- Constructive interference occurs when two identical waves are superimposed in phase.

- Destructive interference occurs when two identical waves are superimposed exactly out of phase.
- A standing wave is one in which two waves superimpose to produce a wave that varies in amplitude but does not propagate.
- Nodes are points of no motion in standing waves.
- An antinode is the location of maximum amplitude of a standing wave.
- Waves on a string are resonant standing waves with a fundamental frequency and can occur at higher multiples of the fundamental, called overtones or harmonics.
- Beats occur when waves of similar frequencies f_1 and f_2 are superimposed. The resulting amplitude oscillates with a beat frequency given by
Equation:

$$f_B = |f_1 - f_2|.$$

Conceptual Questions

Exercise:

Problem:

Speakers in stereo systems have two color-coded terminals to indicate how to hook up the wires. If the wires are reversed, the speaker moves in a direction opposite that of a properly connected speaker. Explain why it is important to have both speakers connected the same way.

Problems & Exercises

Exercise:

Problem:

A car has two horns, one emitting a frequency of 199 Hz and the other emitting a frequency of 203 Hz. What beat frequency do they produce?

Solution:

$$f = 4 \text{ Hz}$$

Exercise:**Problem:**

The middle-C hammer of a piano hits two strings, producing beats of 1.50 Hz. One of the strings is tuned to 260.00 Hz. What frequencies could the other string have?

Exercise:**Problem:**

Two tuning forks having frequencies of 460 and 464 Hz are struck simultaneously. What average frequency will you hear, and what will the beat frequency be?

Solution:

462 Hz,

4 Hz

Exercise:**Problem:**

Twin jet engines on an airplane are producing an average sound frequency of 4100 Hz with a beat frequency of 0.500 Hz. What are their individual frequencies?

Exercise:**Problem:**

A wave traveling on a Slinky® that is stretched to 4 m takes 2.4 s to travel the length of the Slinky and back again. (a) What is the speed of the wave? (b) Using the same Slinky stretched to the same length, a standing wave is created which consists of three antinodes and four nodes. At what frequency must the Slinky be oscillating?

Solution:

(a) 3.33 m/s

(b) 1.25 Hz

Exercise:**Problem:**

Three adjacent keys on a piano (F, F-sharp, and G) are struck simultaneously, producing frequencies of 349, 370, and 392 Hz. What beat frequencies are produced by this discordant combination?

Glossary

antinode

the location of maximum amplitude in standing waves

beat frequency

the frequency of the amplitude fluctuations of a wave

constructive interference

when two waves arrive at the same point exactly in phase; that is, the crests of the two waves are precisely aligned, as are the troughs

destructive interference

when two identical waves arrive at the same point exactly out of phase; that is, precisely aligned crest to trough

fundamental frequency

the lowest frequency of a periodic waveform

nodes

the points where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave

overtones

multiples of the fundamental frequency of a sound

superposition

the phenomenon that occurs when two or more waves arrive at the same point

Energy in Waves: Intensity

- Calculate the intensity and the power of rays and waves.



The destructive effect of an earthquake is palpable evidence of the energy carried in these waves. The Richter scale rating of earthquakes is related to both their amplitude and the energy they carry.

(credit: Petty Officer 2nd Class Candice Villarreal, U.S. Navy)

All waves carry energy. The energy of some waves can be directly observed. Earthquakes can shake whole cities to the ground, performing the work of thousands of wrecking balls.

Loud sounds pulverize nerve cells in the inner ear, causing permanent hearing loss. Ultrasound is used for deep-heat treatment of muscle strains. A laser beam can burn away a malignancy. Water waves chew up beaches.

The amount of energy in a wave is related to its amplitude. Large-amplitude earthquakes produce large ground displacements. Loud sounds have higher pressure amplitudes and come from larger-amplitude source vibrations than

soft sounds. Large ocean breakers churn up the shore more than small ones. More quantitatively, a wave is a displacement that is resisted by a restoring force. The larger the displacement x , the larger the force $F = kx$ needed to create it. Because work W is related to force multiplied by distance (Fx) and energy is put into the wave by the work done to create it, the energy in a wave is related to amplitude. In fact, a wave's energy is directly proportional to its amplitude squared because

Equation:

$$W \propto Fx = kx^2.$$

The energy effects of a wave depend on time as well as amplitude. For example, the longer deep-heat ultrasound is applied, the more energy it transfers. Waves can also be concentrated or spread out. Sunlight, for example, can be focused to burn wood. Earthquakes spread out, so they do less damage the farther they get from the source. In both cases, changing the area the waves cover has important effects. All these pertinent factors are included in the definition of **intensity** I as power per unit area:

Equation:

$$I = \frac{P}{A}$$

where P is the power carried by the wave through area A . The definition of intensity is valid for any energy in transit, including that carried by waves. The SI unit for intensity is watts per square meter (W/m^2). For example, infrared and visible energy from the Sun impinge on Earth at an intensity of $1300 \text{ W}/\text{m}^2$ just above the atmosphere. There are other intensity-related units in use, too. The most common is the decibel. For example, a 90 decibel sound level corresponds to an intensity of $10^{-3} \text{ W}/\text{m}^2$. (This quantity is not much power per unit area considering that 90 decibels is a relatively high sound level. Decibels will be discussed in some detail in a later chapter.

Example:**Calculating intensity and power: How much energy is in a ray of sunlight?**

The average intensity of sunlight on Earth's surface is about 700 W/m^2 .

(a) Calculate the amount of energy that falls on a solar collector having an area of 0.500 m^2 in 4.00 h .

(b) What intensity would such sunlight have if concentrated by a magnifying glass onto an area 200 times smaller than its own?

Strategy a

Because power is energy per unit time or $P = \frac{E}{t}$, the definition of intensity can be written as $I = \frac{P}{A} = \frac{E/t}{A}$, and this equation can be solved for E with the given information.

Solution a

1. Begin with the equation that states the definition of intensity:

Equation:

$$I = \frac{P}{A}.$$

2. Replace P with its equivalent E/t :

Equation:

$$I = \frac{E/t}{A}.$$

3. Solve for E :

Equation:

$$E = IAt.$$

4. Substitute known values into the equation:

Equation:

$$E = (700 \text{ W/m}^2)(0.500 \text{ m}^2)[(4.00 \text{ h})(3600 \text{ s/h})].$$

5. Calculate to find E and convert units:

Equation:

$$5.04 \times 10^6 \text{ J},$$

Discussion a

The energy falling on the solar collector in 4 h in part is enough to be useful—for example, for heating a significant amount of water.

Strategy b

Taking a ratio of new intensity to old intensity and using primes for the new quantities, we will find that it depends on the ratio of the areas. All other quantities will cancel.

Solution b

1. Take the ratio of intensities, which yields:

Equation:

$$\frac{I'}{I} = \frac{P'/A'}{P/A} = \frac{A}{A'} \left(\text{The powers cancel because } P' = P \right).$$

2. Identify the knowns:

Equation:

$$A = 200A',$$

Equation:

$$\frac{I'}{I} = 200.$$

3. Substitute known quantities:

Equation:

$$I' = 200I = 200(700 \text{ W/m}^2).$$

4. Calculate to find I' :

Equation:

$$I' = 1.40 \times 10^5 \text{ W/m}^2.$$

Discussion b

Decreasing the area increases the intensity considerably. The intensity of the concentrated sunlight could even start a fire.

Example:**Determine the combined intensity of two waves: Perfect constructive interference**

If two identical waves, each having an intensity of 1.00 W/m^2 , interfere perfectly constructively, what is the intensity of the resulting wave?

Strategy

We know from [Superposition and Interference](#) that when two identical waves, which have equal amplitudes X , interfere perfectly constructively, the resulting wave has an amplitude of $2X$. Because a wave's intensity is proportional to amplitude squared, the intensity of the resulting wave is four times as great as in the individual waves.

Solution

1. Recall that intensity is proportional to amplitude squared.
2. Calculate the new amplitude:

Equation:

$$I' \propto (X')^2 = (2X)^2 = 4X^2.$$

3. Recall that the intensity of the old amplitude was:

Equation:

$$I \propto X^2.$$

4. Take the ratio of new intensity to the old intensity. This gives:

Equation:

$$\frac{I'}{I} = 4.$$

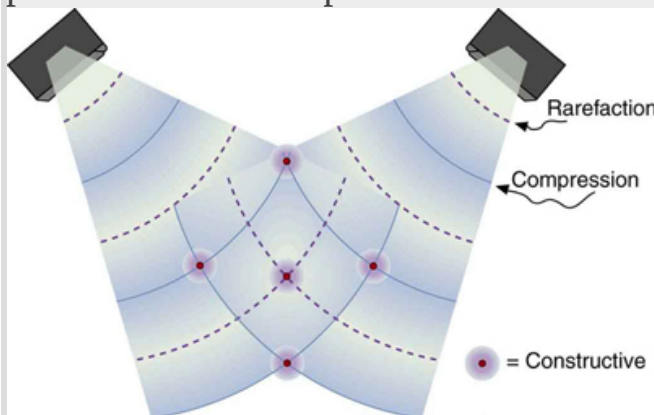
5. Calculate to find I' :

Equation:

$$I' = 4I = 4.00 \text{ W/m}^2.$$

Discussion

The intensity goes up by a factor of 4 when the amplitude doubles. This answer is a little disquieting. The two individual waves each have intensities of 1.00 W/m^2 , yet their sum has an intensity of 4.00 W/m^2 , which may appear to violate conservation of energy. This violation, of course, cannot happen. What does happen is intriguing. The area over which the intensity is 4.00 W/m^2 is much less than the area covered by the two waves before they interfered. There are other areas where the intensity is zero. The addition of waves is not as simple as our first look in [Superposition and Interference](#) suggested. We actually get a pattern of both constructive interference and destructive interference whenever two waves are added. For example, if we have two stereo speakers putting out 1.00 W/m^2 each, there will be places in the room where the intensity is 4.00 W/m^2 , other places where the intensity is zero, and others in between. [\[link\]](#) shows what this interference might look like. We will pursue interference patterns elsewhere in this text.



These stereo speakers produce both constructive interference and destructive interference in the room, a property common to the

superposition of all types of waves.
The shading is proportional to
intensity.

Exercise:

Check Your Understanding

Problem:

Which measurement of a wave is most important when determining the wave's intensity?

Solution:

Amplitude, because a wave's energy is directly proportional to its amplitude squared.

Section Summary

Intensity is defined to be the power per unit area:

$$I = \frac{P}{A} \text{ and has units of } \text{W}/\text{m}^2.$$

Conceptual Questions

Exercise:

Problem:

Two identical waves undergo pure constructive interference. Is the resultant intensity twice that of the individual waves? Explain your answer.

Exercise:

Problem:

Circular water waves decrease in amplitude as they move away from where a rock is dropped. Explain why.

Problems & Exercises**Exercise:****Problem: Medical Application**

Ultrasound of intensity $1.50 \times 10^2 \text{ W/m}^2$ is produced by the rectangular head of a medical imaging device measuring 3.00 by 5.00 cm. What is its power output?

Solution:

0.225 W

Exercise:**Problem:**

The low-frequency speaker of a stereo set has a surface area of 0.05 m^2 and produces 1W of acoustical power. What is the intensity at the speaker? If the speaker projects sound uniformly in all directions, at what distance from the speaker is the intensity 0.1 W/m^2 ?

Exercise:**Problem:**

To increase intensity of a wave by a factor of 50, by what factor should the amplitude be increased?

Solution:

7.07

Exercise:**Problem: Engineering Application**

A device called an insolation meter is used to measure the intensity of sunlight has an area of 100 cm^2 and registers 6.50 W . What is the intensity in W/m^2 ?

Exercise:**Problem: Astronomy Application**

Energy from the Sun arrives at the top of the Earth's atmosphere with an intensity of 1.30 kW/m^2 . How long does it take for $1.8 \times 10^9 \text{ J}$ to arrive on an area of 1.00 m^2 ?

Solution:

16.0 d

Exercise:**Problem:**

Suppose you have a device that extracts energy from ocean breakers in direct proportion to their intensity. If the device produces 10.0 kW of power on a day when the breakers are 1.20 m high, how much will it produce when they are 0.600 m high?

Solution:

2.50 kW

Exercise:**Problem: Engineering Application**

(a) A photovoltaic array of (solar cells) is 10.0% efficient in gathering solar energy and converting it to electricity. If the average intensity of

sunlight on one day is 700 W/m^2 , what area should your array have to gather energy at the rate of 100 W ? (b) What is the maximum cost of the array if it must pay for itself in two years of operation averaging 10.0 hours per day? Assume that it earns money at the rate of 9.00 ¢ per kilowatt-hour.

Exercise:

Problem:

A microphone receiving a pure sound tone feeds an oscilloscope, producing a wave on its screen. If the sound intensity is originally $2.00 \times 10^{-5} \text{ W/m}^2$, but is turned up until the amplitude increases by 30.0% , what is the new intensity?

Solution:

$$3.38 \times 10^{-5} \text{ W/m}^2$$

Exercise:

Problem: Medical Application

(a) What is the intensity in W/m^2 of a laser beam used to burn away cancerous tissue that, when 90.0% absorbed, puts 500 J of energy into a circular spot 2.00 mm in diameter in 4.00 s ? (b) Discuss how this intensity compares to the average intensity of sunlight (about 700 W/m^2) and the implications that would have if the laser beam entered your eye. Note how your answer depends on the time duration of the exposure.

Glossary

intensity

power per unit area

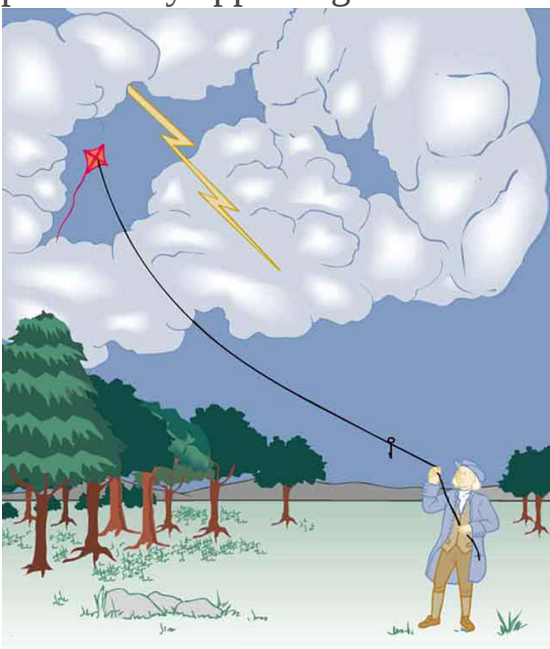
Introduction to Electric Charge and Electric Field

class="introduction"

Static electricity from this plastic slide causes the child's hair to stand on end. The sliding motion stripped electrons away from the child's body, leaving an excess of positive charges, which repel each other along each strand of hair. (credit: Ken Bosma/Wikimedia Commons)



The image of American politician and scientist Benjamin Franklin (1706–1790) flying a kite in a thunderstorm is familiar to every schoolchild. (See [\[link\]](#).) In this experiment, Franklin demonstrated a connection between lightning and **static electricity**. Sparks were drawn from a key hung on a kite string during an electrical storm. These sparks were like those produced by static electricity, such as the spark that jumps from your finger to a metal doorknob after you walk across a wool carpet. What Franklin demonstrated in his dangerous experiment was a connection between phenomena on two different scales: one the grand power of an electrical storm, the other an effect of more human proportions. Connections like this one reveal the underlying unity of the laws of nature, an aspect we humans find particularly appealing.



When Benjamin Franklin demonstrated that lightning was related to static electricity, he made a connection that is now part of the evidence that all directly experienced forces except the gravitational force are manifestations of the electromagnetic force.

Much has been written about Franklin. His experiments were only part of the life of a man who was a scientist, inventor, revolutionary, statesman, and writer. Franklin's experiments were not performed in isolation, nor were they the only ones to reveal connections.

For example, the Italian scientist Luigi Galvani (1737–1798) performed a series of experiments in which static electricity was used to stimulate contractions of leg muscles of dead frogs, an effect already known in humans subjected to static discharges. But Galvani also found that if he joined two metal wires (say copper and zinc) end to end and touched the other ends to muscles, he produced the same effect in frogs as static discharge. Alessandro Volta (1745–1827), partly inspired by Galvani's work, experimented with various combinations of metals and developed the battery.

During the same era, other scientists made progress in discovering fundamental connections. The periodic table was developed as the systematic properties of the elements were discovered. This influenced the development and refinement of the concept of atoms as the basis of matter. Such submicroscopic descriptions of matter also help explain a great deal more.

Atomic and molecular interactions, such as the forces of friction, cohesion, and adhesion, are now known to be manifestations of the **electromagnetic force**. Static electricity is just one aspect of the electromagnetic force, which also includes moving electricity and magnetism.

All the macroscopic forces that we experience directly, such as the sensations of touch and the tension in a rope, are due to the electromagnetic force, one of the four fundamental forces in nature. The gravitational force, another fundamental force, is actually sensed through the electromagnetic interaction of molecules, such as between those in our feet and those on the top of a bathroom scale. (The other two fundamental forces, the strong nuclear force and the weak nuclear force, cannot be sensed on the human scale.)

This chapter begins the study of electromagnetic phenomena at a fundamental level. The next several chapters will cover static electricity, moving electricity, and magnetism—collectively known as electromagnetism. In this chapter, we begin with the study of electric phenomena due to charges that are at least temporarily stationary, called electrostatics, or static electricity.

Glossary

static electricity

a buildup of electric charge on the surface of an object

electromagnetic force

one of the four fundamental forces of nature; the electromagnetic force consists of static electricity, moving electricity and magnetism

Static Electricity and Charge: Conservation of Charge

- Define electric charge, and describe how the two types of charge interact.
- Describe three common situations that generate static electricity.
- State the law of conservation of charge.



Borneo amber was mined in Sabah, Malaysia, from shale-sandstone-mudstone veins.

When a piece of amber is rubbed with a piece of silk, the amber gains more electrons, giving it a net negative charge.

At the same time, the silk, having lost electrons, becomes positively charged. (credit: Sebakoamber, Wikimedia Commons)

What makes plastic wrap cling? Static electricity. Not only are applications of static electricity common these days, its existence has been known since ancient times. The first record of its effects dates to ancient Greeks who noted more than 500 years B.C. that polishing amber temporarily enabled it

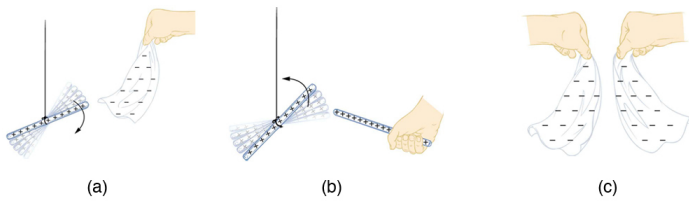
to attract bits of straw (see [\[link\]](#)). The very word *electric* derives from the Greek word for amber (*electron*).

Many of the characteristics of static electricity can be explored by rubbing things together. Rubbing creates the spark you get from walking across a wool carpet, for example. Static cling generated in a clothes dryer and the attraction of straw to recently polished amber also result from rubbing. Similarly, lightning results from air movements under certain weather conditions. You can also rub a balloon on your hair, and the static electricity created can then make the balloon cling to a wall. We also have to be cautious of static electricity, especially in dry climates. When we pump gasoline, we are warned to discharge ourselves (after sliding across the seat) on a metal surface before grabbing the gas nozzle. Attendants in hospital operating rooms must wear booties with aluminum foil on the bottoms to avoid creating sparks which may ignite the oxygen being used.

Some of the most basic characteristics of static electricity include:

- The effects of static electricity are explained by a physical quantity not previously introduced, called electric charge.
- There are only two types of charge, one called positive and the other called negative.
- Like charges repel, whereas unlike charges attract.
- The force between charges decreases with distance.

How do we know there are two types of **electric charge**? When various materials are rubbed together in controlled ways, certain combinations of materials always produce one type of charge on one material and the opposite type on the other. By convention, we call one type of charge “positive”, and the other type “negative.” For example, when glass is rubbed with silk, the glass becomes positively charged and the silk negatively charged. Since the glass and silk have opposite charges, they attract one another like clothes that have rubbed together in a dryer. Two glass rods rubbed with silk in this manner will repel one another, since each rod has positive charge on it. Similarly, two silk cloths so rubbed will repel, since both cloths have negative charge. [\[link\]](#) shows how these simple materials can be used to explore the nature of the force between charges.



A glass rod becomes positively charged when rubbed with silk, while the silk becomes negatively charged.

(a) The glass rod is attracted to the silk because their charges are opposite. (b) Two similarly charged glass rods repel. (c) Two similarly charged silk cloths repel.

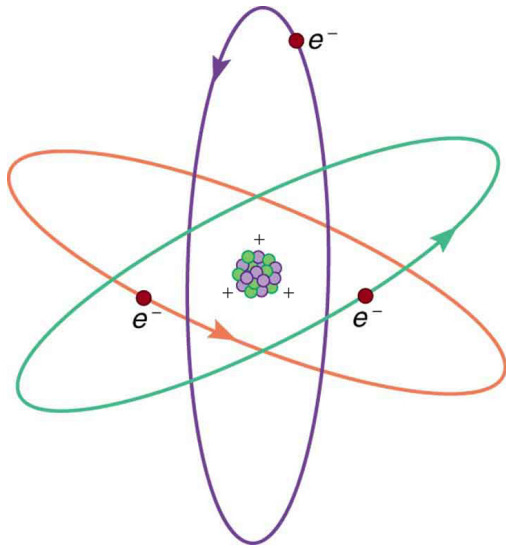
More sophisticated questions arise. Where do these charges come from? Can you create or destroy charge? Is there a smallest unit of charge? Exactly how does the force depend on the amount of charge and the distance between charges? Such questions obviously occurred to Benjamin Franklin and other early researchers, and they interest us even today.

Charge Carried by Electrons and Protons

Franklin wrote in his letters and books that he could see the effects of electric charge but did not understand what caused the phenomenon. Today we have the advantage of knowing that normal matter is made of atoms, and that atoms contain positive and negative charges, usually in equal amounts.

[\[link\]](#) shows a simple model of an atom with negative **electrons** orbiting its positive nucleus. The nucleus is positive due to the presence of positively charged **protons**. Nearly all charge in nature is due to electrons and protons, which are two of the three building blocks of most matter. (The third is the neutron, which is neutral, carrying no charge.) Other charge-carrying particles are observed in cosmic rays and nuclear decay, and are created in

particle accelerators. All but the electron and proton survive only a short time and are quite rare by comparison.



This simplified (and not to scale) view of an atom is called the planetary model of the atom.

Negative electrons orbit a much heavier positive nucleus, as the planets orbit the much heavier sun. There the similarity ends, because forces in the atom are electromagnetic, whereas those in the planetary system are gravitational.

Normal macroscopic amounts of matter contain immense numbers of atoms and molecules and, hence, even greater numbers of individual

negative and positive charges.

The charges of electrons and protons are identical in magnitude but opposite in sign. Furthermore, all charged objects in nature are integral multiples of this basic quantity of charge, meaning that all charges are made of combinations of a basic unit of charge. Usually, charges are formed by combinations of electrons and protons. The magnitude of this basic charge is

Equation:

$$|q_e| = 1.60 \times 10^{-19} \text{ C}.$$

The symbol q is commonly used for charge and the subscript e indicates the charge of a single electron (or proton).

The SI unit of charge is the coulomb (C). The number of protons needed to make a charge of 1.00 C is

Equation:

$$1.00 \text{ C} \times \frac{1 \text{ proton}}{1.60 \times 10^{-19} \text{ C}} = 6.25 \times 10^{18} \text{ protons}.$$

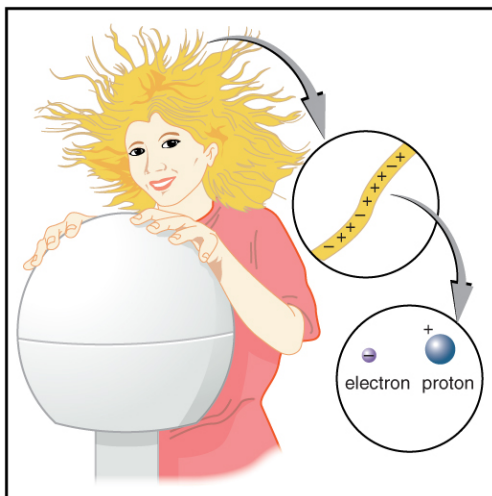
Similarly, 6.25×10^{18} electrons have a combined charge of -1.00 coulomb. Just as there is a smallest bit of an element (an atom), there is a smallest bit of charge. There is no directly observed charge smaller than $|q_e|$ (see [Things Great and Small: The Submicroscopic Origin of Charge](#)), and all observed charges are integral multiples of $|q_e|$.

Note:

Things Great and Small: The Submicroscopic Origin of Charge

With the exception of exotic, short-lived particles, all charge in nature is carried by electrons and protons. Electrons carry the charge we have named negative. Protons carry an equal-magnitude charge that we call positive. (See [\[link\]](#).) Electron and proton charges are considered fundamental building blocks, since all other charges are integral multiples of those carried by electrons and protons. Electrons and protons are also two of the three fundamental building blocks of ordinary matter. The neutron is the third and has zero total charge.

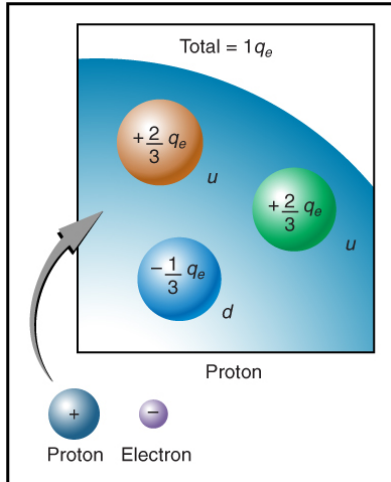
[\[link\]](#) shows a person touching a Van de Graaff generator and receiving excess positive charge. The expanded view of a hair shows the existence of both types of charges but an excess of positive. The repulsion of these positive like charges causes the strands of hair to repel other strands of hair and to stand up. The further blowup shows an artist's conception of an electron and a proton perhaps found in an atom in a strand of hair.



When this person touches
a Van de Graaff
generator, she receives an
excess of positive charge,
causing her hair to stand
on end. The charges in

one hair are shown. An artist's conception of an electron and a proton illustrate the particles carrying the negative and positive charges. We cannot really see these particles with visible light because they are so small (the electron seems to be an infinitesimal point), but we know a great deal about their measurable properties, such as the charges they carry.

The electron seems to have no substructure; in contrast, when the substructure of protons is explored by scattering extremely energetic electrons from them, it appears that there are point-like particles inside the proton. These sub-particles, named quarks, have never been directly observed, but they are believed to carry fractional charges as seen in [\[link\]](#). Charges on electrons and protons and all other directly observable particles are unitary, but these quark substructures carry charges of either $-\frac{1}{3}$ or $+\frac{2}{3}$. There are continuing attempts to observe fractional charge directly and to learn of the properties of quarks, which are perhaps the ultimate substructure of matter.



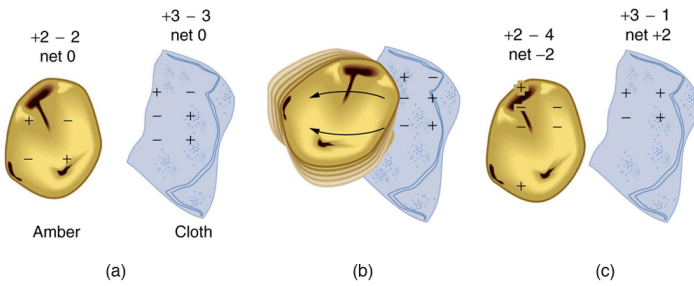
Artist's conception of fractional quark charges inside a proton. A group of three quark charges add up to the single positive charge on the proton:

$$-\frac{1}{3}q_e + \frac{2}{3}q_e + \frac{2}{3}q_e = +1q_e$$

.

Separation of Charge in Atoms

Charges in atoms and molecules can be separated—for example, by rubbing materials together. Some atoms and molecules have a greater affinity for electrons than others and will become negatively charged by close contact in rubbing, leaving the other material positively charged. (See [link](#).) Positive charge can similarly be induced by rubbing. Methods other than rubbing can also separate charges. Batteries, for example, use combinations of substances that interact in such a way as to separate charges. Chemical interactions may transfer negative charge from one substance to the other, making one battery terminal negative and leaving the first one positive.



When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, and only a few of them are shown here. (b) When rubbed together, some negative charge is transferred to the amber, leaving the cloth with a net positive charge. (c) When separated, the amber and cloth now have net charges, but the absolute value of the net positive and negative charges will be equal.

No charge is actually created or destroyed when charges are separated as we have been discussing. Rather, existing charges are moved about. In fact, in all situations the total amount of charge is always constant. This universally obeyed law of nature is called the **law of conservation of charge**.

Note:

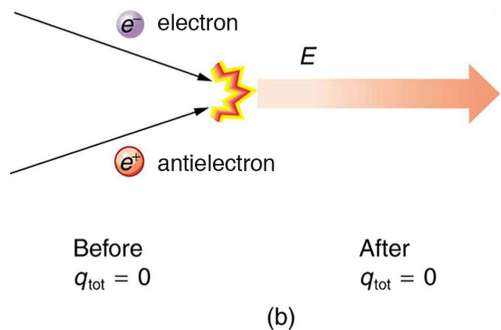
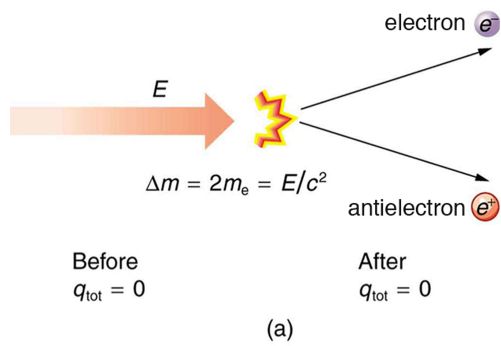
Law of Conservation of Charge

Total charge is constant in any process.

In more exotic situations, such as in particle accelerators, mass, Δm , can be created from energy in the amount $\Delta m = \frac{E}{c^2}$. Sometimes, the created mass is charged, such as when an electron is created. Whenever a charged particle is created, another having an opposite charge is always created along with it, so that the total charge created is zero. Usually, the two particles are “matter-antimatter” counterparts. For example, an antielectron would usually be created at the same time as an electron. The antielectron has a positive charge (it is called a positron), and so the total charge created is zero. (See [\[link\]](#).) All particles have antimatter counterparts with opposite signs. When matter and antimatter counterparts are brought together, they completely annihilate one another. By annihilate, we mean that the mass of the two particles is converted to energy E , again obeying the relationship $\Delta m = \frac{E}{c^2}$. Since the two particles have equal and opposite charge, the total charge is zero before and after the annihilation; thus, total charge is conserved.

Note:**Making Connections: Conservation Laws**

Only a limited number of physical quantities are universally conserved. Charge is one—energy, momentum, and angular momentum are others. Because they are conserved, these physical quantities are used to explain more phenomena and form more connections than other, less basic quantities. We find that conserved quantities give us great insight into the rules followed by nature and hints to the organization of nature. Discoveries of conservation laws have led to further discoveries, such as the weak nuclear force and the quark substructure of protons and other particles.



(a) When enough energy is present, it can be converted into matter. Here the matter created is an electron–antielectron pair. (m_e is the electron’s mass.) The total charge before and after this event is zero. (b) When matter and antimatter collide, they annihilate each other; the total charge is conserved at zero before and after the annihilation.

The law of conservation of charge is absolute—it has never been observed to be violated. Charge, then, is a special physical quantity, joining a very

short list of other quantities in nature that are always conserved. Other conserved quantities include energy, momentum, and angular momentum.

Note:

PhET Explorations: Balloons and Static Electricity

Why does a balloon stick to your sweater? Rub a balloon on a sweater, then let go of the balloon and it flies over and sticks to the sweater. View the charges in the sweater, balloons, and the wall.

https://phet.colorado.edu/sims/html/balloons-and-static-electricity/latest/balloons-and-static-electricity_en.html

Section Summary

- There are only two types of charge, which we call positive and negative.
- Like charges repel, unlike charges attract, and the force between charges decreases with the square of the distance.
- The vast majority of positive charge in nature is carried by protons, while the vast majority of negative charge is carried by electrons.
- The electric charge of one electron is equal in magnitude and opposite in sign to the charge of one proton.
- An ion is an atom or molecule that has nonzero total charge due to having unequal numbers of electrons and protons.
- The SI unit for charge is the coulomb (C), with protons and electrons having charges of opposite sign but equal magnitude; the magnitude of this basic charge $|q_e|$ is

Equation:

$$|q_e| = 1.60 \times 10^{-19} \text{ C}.$$

- Whenever charge is created or destroyed, equal amounts of positive and negative are involved.
- Most often, existing charges are separated from neutral objects to obtain some net charge.

- Both positive and negative charges exist in neutral objects and can be separated by rubbing one object with another. For macroscopic objects, negatively charged means an excess of electrons and positively charged means a depletion of electrons.
- The law of conservation of charge ensures that whenever a charge is created, an equal charge of the opposite sign is created at the same time.

Conceptual Questions

Exercise:

Problem:

There are very large numbers of charged particles in most objects. Why, then, don't most objects exhibit static electricity?

Exercise:

Problem:

Why do most objects tend to contain nearly equal numbers of positive and negative charges?

Problems & Exercises

Exercise:

Problem:

Common static electricity involves charges ranging from nanocoulombs to microcoulombs. (a) How many electrons are needed to form a charge of -2.00 nC (b) How many electrons must be removed from a neutral object to leave a net charge of $0.500 \mu\text{C}$?

Solution:

(a) 1.25×10^{10}

(b) 3.13×10^{12}

Exercise:

Problem:

If 1.80×10^{20} electrons move through a pocket calculator during a full day's operation, how many coulombs of charge moved through it?

Exercise:

Problem:

To start a car engine, the car battery moves 3.75×10^{21} electrons through the starter motor. How many coulombs of charge were moved?

Solution:

-600 C

Exercise:

Problem:

A certain lightning bolt moves 40.0 C of charge. How many fundamental units of charge $|q_e|$ is this?

Glossary

electric charge

a physical property of an object that causes it to be attracted toward or repelled from another charged object; each charged object generates and is influenced by a force called an electromagnetic force

law of conservation of charge

states that whenever a charge is created, an equal amount of charge with the opposite sign is created simultaneously

electron

a particle orbiting the nucleus of an atom and carrying the smallest unit of negative charge

proton

a particle in the nucleus of an atom and carrying a positive charge equal in magnitude and opposite in sign to the amount of negative charge carried by an electron

Conductors and Insulators

- Define conductor and insulator, explain the difference, and give examples of each.
- Describe three methods for charging an object.
- Explain what happens to an electric force as you move farther from the source.
- Define polarization.

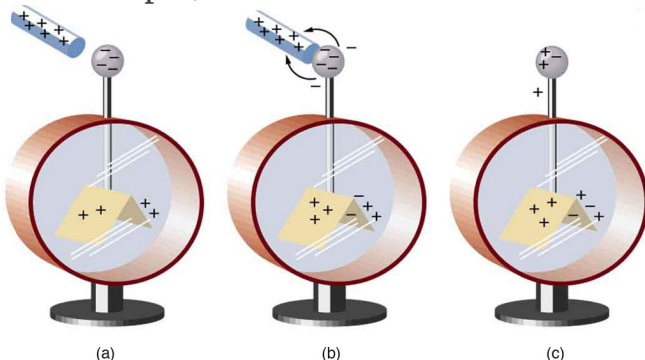


This power adapter uses metal wires and connectors to conduct electricity from the wall socket to a laptop computer. The conducting wires allow electrons to move freely through the cables, which are shielded by rubber and plastic. These materials act as insulators that don't allow electric charge to escape outward. (credit: Evan-Amos, Wikimedia Commons)

Some substances, such as metals and salty water, allow charges to move through them with relative ease. Some of the electrons in metals and similar conductors are not bound to individual atoms or sites in the material. These **free electrons** can move through the material much as air moves through loose sand. Any substance that has free electrons and allows charge to move

relatively freely through it is called a **conductor**. The moving electrons may collide with fixed atoms and molecules, losing some energy, but they can move in a conductor. Superconductors allow the movement of charge without any loss of energy. Salty water and other similar conducting materials contain free ions that can move through them. An ion is an atom or molecule having a positive or negative (nonzero) total charge. In other words, the total number of electrons is not equal to the total number of protons.

Other substances, such as glass, do not allow charges to move through them. These are called **insulators**. Electrons and ions in insulators are bound in the structure and cannot move easily—as much as 10^{23} times more slowly than in conductors. Pure water and dry table salt are insulators, for example, whereas molten salt and salty water are conductors.



An electroscope is a favorite instrument in physics demonstrations and student laboratories. It is typically made with gold foil leaves hung from a (conducting) metal stem and is insulated from the room air in a glass-walled container. (a) A positively charged glass rod is brought near the tip of the electroscope, attracting electrons to the top and leaving a net positive charge on the leaves. Like charges in the light flexible gold leaves

repel, separating them. (b) When the rod is touched against the ball, electrons are attracted and transferred, reducing the net charge on the glass rod but leaving the electroscope positively charged. (c) The excess charges are evenly distributed in the stem and leaves of the electroscope once the glass rod is removed.

Charging by Contact

[\[link\]](#) shows an electroscope being charged by touching it with a positively charged glass rod. Because the glass rod is an insulator, it must actually touch the electroscope to transfer charge to or from it. (Note that the extra positive charges reside on the surface of the glass rod as a result of rubbing it with silk before starting the experiment.) Since only electrons move in metals, we see that they are attracted to the top of the electroscope. There, some are transferred to the positive rod by touch, leaving the electroscope with a net positive charge.

Electrostatic repulsion in the leaves of the charged electroscope separates them. The electrostatic force has a horizontal component that results in the leaves moving apart as well as a vertical component that is balanced by the gravitational force. Similarly, the electroscope can be negatively charged by contact with a negatively charged object.

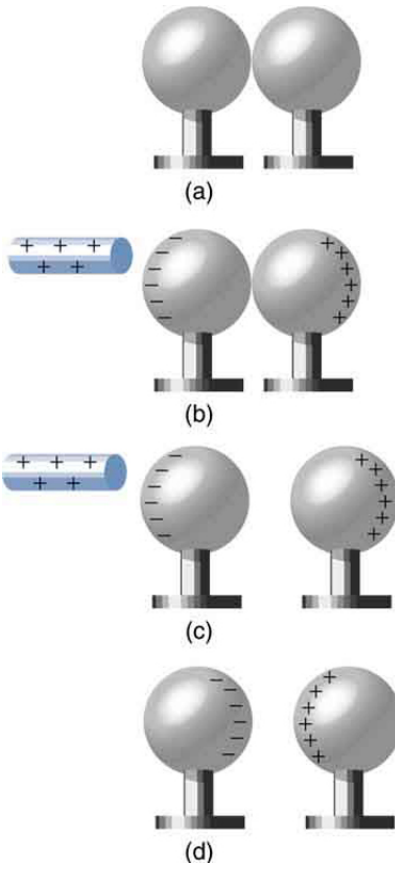
Charging by Induction

It is not necessary to transfer excess charge directly to an object in order to charge it. [\[link\]](#) shows a method of **induction** wherein a charge is created in a nearby object, without direct contact. Here we see two neutral metal spheres in contact with one another but insulated from the rest of the world.

A positively charged rod is brought near one of them, attracting negative charge to that side, leaving the other sphere positively charged.

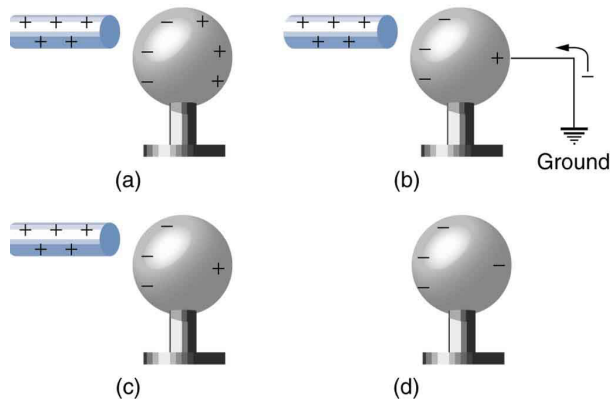
This is an example of induced **polarization** of neutral objects. Polarization is the separation of charges in an object that remains neutral. If the spheres are now separated (before the rod is pulled away), each sphere will have a net charge. Note that the object closest to the charged rod receives an opposite charge when charged by induction. Note also that no charge is removed from the charged rod, so that this process can be repeated without depleting the supply of excess charge.

Another method of charging by induction is shown in [\[link\]](#). The neutral metal sphere is polarized when a charged rod is brought near it. The sphere is then grounded, meaning that a conducting wire is run from the sphere to the ground. Since the earth is large and most ground is a good conductor, it can supply or accept excess charge easily. In this case, electrons are attracted to the sphere through a wire called the ground wire, because it supplies a conducting path to the ground. The ground connection is broken before the charged rod is removed, leaving the sphere with an excess charge opposite to that of the rod. Again, an opposite charge is achieved when charging by induction and the charged rod loses none of its excess charge.



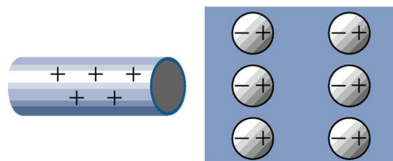
Charging by induction. (a) Two uncharged or neutral metal spheres are in contact with each other but insulated from the rest of the world. (b) A positively charged glass rod is brought near the sphere on the left, attracting negative charge and leaving the other sphere positively charged. (c) The

spheres are separated before the rod is removed, thus separating negative and positive charge. (d) The spheres retain net charges after the inducing rod is removed—without ever having been touched by a charged object.

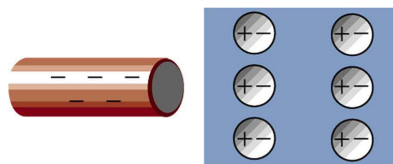


Charging by induction, using a ground connection. (a) A positively charged rod is brought near a neutral metal sphere, polarizing it. (b) The sphere is grounded, allowing electrons to be attracted from the earth's ample supply. (c) The ground connection is broken. (d) The positive rod is

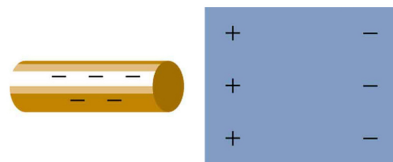
removed, leaving the sphere
with an induced negative
charge.



(a)



(b)



(c)

Both positive and
negative objects
attract a neutral
object by polarizing
its molecules. (a) A
positive object
brought near a
neutral insulator
polarizes its
molecules. There is
a slight shift in the
distribution of the
electrons orbiting
the molecule, with

unlike charges being brought nearer and like charges moved away. Since the electrostatic force decreases with distance, there is a net attraction. (b) A negative object produces the opposite polarization, but again attracts the neutral object. (c) The same effect occurs for a conductor; since the unlike charges are closer, there is a net attraction.

Neutral objects can be attracted to any charged object. The pieces of straw attracted to polished amber are neutral, for example. If you run a plastic comb through your hair, the charged comb can pick up neutral pieces of paper. [\[link\]](#) shows how the polarization of atoms and molecules in neutral objects results in their attraction to a charged object.

When a charged rod is brought near a neutral substance, an insulator in this case, the distribution of charge in atoms and molecules is shifted slightly. Opposite charge is attracted nearer the external charged rod, while like charge is repelled. Since the electrostatic force decreases with distance, the repulsion of like charges is weaker than the attraction of unlike charges, and so there is a net attraction. Thus a positively charged glass rod attracts neutral pieces of paper, as will a negatively charged rubber rod. Some

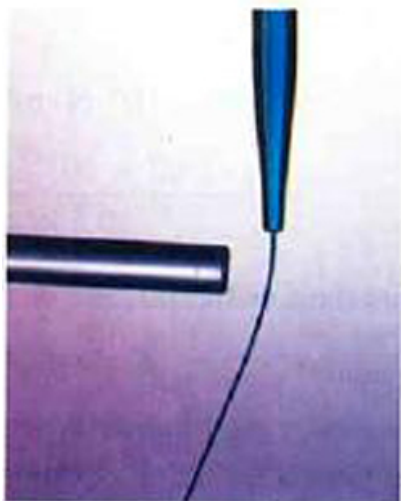
molecules, like water, are polar molecules. Polar molecules have a natural or inherent separation of charge, although they are neutral overall. Polar molecules are particularly affected by other charged objects and show greater polarization effects than molecules with naturally uniform charge distributions.

Exercise:

Check Your Understanding

Problem:

Can you explain the attraction of water to the charged rod in the figure below?



Solution:

Answer

Water molecules are polarized, giving them slightly positive and slightly negative sides. This makes water even more susceptible to a charged rod's attraction. As the water flows downward, due to the force of gravity, the charged conductor exerts a net attraction to the opposite charges in the stream of water, pulling it closer.

Note:

PhET Explorations: John Travoltage

Make sparks fly with John Travoltage. Wiggle Johnnie's foot and he picks up charges from the carpet. Bring his hand close to the door knob and get rid of the excess charge.

https://phet.colorado.edu/sims/html/john-travoltage/latest/john-travoltage_en.html

Section Summary

- Polarization is the separation of positive and negative charges in a neutral object.
- A conductor is a substance that allows charge to flow freely through its atomic structure.
- An insulator holds charge within its atomic structure.
- Objects with like charges repel each other, while those with unlike charges attract each other.
- A conducting object is said to be grounded if it is connected to the Earth through a conductor. Grounding allows transfer of charge to and from the earth's large reservoir.
- Objects can be charged by contact with another charged object and obtain the same sign charge.
- If an object is temporarily grounded, it can be charged by induction, and obtains the opposite sign charge.
- Polarized objects have their positive and negative charges concentrated in different areas, giving them a non-symmetrical charge.
- Polar molecules have an inherent separation of charge.

Conceptual Questions

Exercise:

Problem:

An eccentric inventor attempts to levitate by first placing a large negative charge on himself and then putting a large positive charge on the ceiling of his workshop. Instead, while attempting to place a large negative charge on himself, his clothes fly off. Explain.

Exercise:**Problem:**

If you have charged an electroscope by contact with a positively charged object, describe how you could use it to determine the charge of other objects. Specifically, what would the leaves of the electroscope do if other charged objects were brought near its knob?

Exercise:**Problem:**

When a glass rod is rubbed with silk, it becomes positive and the silk becomes negative—yet both attract dust. Does the dust have a third type of charge that is attracted to both positive and negative? Explain.

Exercise:**Problem:**

Why does a car always attract dust right after it is polished? (Note that car wax and car tires are insulators.)

Exercise:**Problem:**

Describe how a positively charged object can be used to give another object a negative charge. What is the name of this process?

Exercise:**Problem:**

What is grounding? What effect does it have on a charged conductor? On a charged insulator?

Problems & Exercises**Exercise:**

Problem:

Suppose a speck of dust in an electrostatic precipitator has 1.0000×10^{12} protons in it and has a net charge of -5.00 nC (a very large charge for a small speck). How many electrons does it have?

Solution:

$$1.03 \times 10^{12}$$

Exercise:**Problem:**

An amoeba has 1.00×10^{16} protons and a net charge of 0.300 pC . (a) How many fewer electrons are there than protons? (b) If you paired them up, what fraction of the protons would have no electrons?

Exercise:**Problem:**

A 50.0 g ball of copper has a net charge of $2.00 \mu\text{C}$. What fraction of the copper's electrons has been removed? (Each copper atom has 29 protons, and copper has an atomic mass of 63.5.)

Solution:

$$9.09 \times 10^{-13}$$

Exercise:**Problem:**

What net charge would you place on a 100 g piece of sulfur if you put an extra electron on $1 \text{ in } 10^{12}$ of its atoms? (Sulfur has an atomic mass of 32.1.)

Exercise:

Problem:

How many coulombs of positive charge are there in 4.00 kg of plutonium, given its atomic mass is 244 and that each plutonium atom has 94 protons?

Solution:

$$1.48 \times 10^8 \text{ C}$$

Glossary

free electron

an electron that is free to move away from its atomic orbit

conductor

a material that allows electrons to move separately from their atomic orbits

insulator

a material that holds electrons securely within their atomic orbits

grounded

when a conductor is connected to the Earth, allowing charge to freely flow to and from Earth's unlimited reservoir

induction

the process by which an electrically charged object brought near a neutral object creates a charge in that object

polarization

slight shifting of positive and negative charges to opposite sides of an atom or molecule

electrostatic repulsion

the phenomenon of two objects with like charges repelling each other

Coulomb's Law

- State Coulomb's law in terms of how the electrostatic force changes with the distance between two objects.
- Calculate the electrostatic force between two charged point forces, such as electrons or protons.
- Compare the electrostatic force to the gravitational attraction for a proton and an electron; for a human and the Earth.



This NASA image of Arp 87 shows the result of a strong gravitational attraction between two galaxies. In contrast, at the subatomic level, the electrostatic attraction between two objects, such as an electron and a proton, is far greater than their mutual attraction due to gravity. (credit: NASA/HST)

Through the work of scientists in the late 18th century, the main features of the **electrostatic force**—the existence of two types of charge, the observation that like charges repel, unlike charges attract, and the decrease of force with distance—were eventually refined, and expressed as a mathematical formula. The mathematical formula for the electrostatic force is called **Coulomb's law** after the French physicist Charles Coulomb (1736–1806), who performed experiments and first proposed a formula to calculate it.

Note:

Coulomb's Law

Equation:

$$F = k \frac{|q_1 q_2|}{r^2}.$$

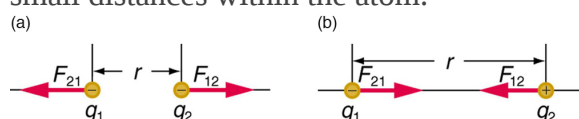
Coulomb's law calculates the magnitude of the force F between two point charges, q_1 and q_2 , separated by a distance r . In SI units, the constant k is equal to

Equation:

$$k = 8.988 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \approx 8.99 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}.$$

The electrostatic force is a vector quantity and is expressed in units of newtons. The force is understood to be along the line joining the two charges. (See [\[link\]](#).)

Although the formula for Coulomb's law is simple, it was no mean task to prove it. The experiments Coulomb did, with the primitive equipment then available, were difficult. Modern experiments have verified Coulomb's law to great precision. For example, it has been shown that the force is inversely proportional to distance between two objects squared ($F \propto 1/r^2$) to an accuracy of 1 part in 10^{16} . No exceptions have ever been found, even at the small distances within the atom.



The magnitude of the electrostatic force F between point charges q_1 and q_2 separated by a distance r is given

by Coulomb's law. Note that

Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force on q_1 is equal in magnitude and opposite in direction to the force it exerts on q_2 .

(a) Like charges. (b) Unlike charges.

Example:

How Strong is the Coulomb Force Relative to the Gravitational Force?

Compare the electrostatic force between an electron and proton separated by 0.530×10^{-10} m with the gravitational force between them. This distance is their average separation in a hydrogen atom.

Strategy

To compare the two forces, we first compute the electrostatic force using Coulomb's law, $F = k \frac{|q_1 q_2|}{r^2}$. We then calculate the gravitational force using Newton's universal law of

gravitation. Finally, we take a ratio to see how the forces compare in magnitude.

Solution

Entering the given and known information about the charges and separation of the electron and proton into the expression of Coulomb's law yields

Equation:

$$F = k \frac{|q_1 q_2|}{r^2}$$

Equation:

$$= (8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \times \frac{(1.60 \times 10^{-19} \text{ C})(1.60 \times 10^{-19} \text{ C})}{(0.530 \times 10^{-10} \text{ m})^2}$$

Thus the Coulomb force is

Equation:

$$F = 8.19 \times 10^{-8} \text{ N}.$$

The charges are opposite in sign, so this is an attractive force. This is a very large force for an electron—it would cause an acceleration of $8.99 \times 10^{22} \text{ m/s}^2$ (verification is left as an end-of-section problem). The gravitational force is given by Newton's law of gravitation as:

Equation:

$$F_G = G \frac{mM}{r^2},$$

where $G = 6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$. Here m and M represent the electron and proton masses, which can be found in the appendices. Entering values for the knowns yields

Equation:

$$F_G = (6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2) \times \frac{(9.11 \times 10^{-31} \text{ kg})(1.67 \times 10^{-27} \text{ kg})}{(0.530 \times 10^{-10} \text{ m})^2} = 3.61 \times 10^{-47} \text{ N}$$

This is also an attractive force, although it is traditionally shown as positive since gravitational force is always attractive. The ratio of the magnitude of the electrostatic force to gravitational force in this case is, thus,

Equation:

$$\frac{F}{F_G} = 2.27 \times 10^{39}.$$

Discussion

This is a remarkably large ratio! Note that this will be the ratio of electrostatic force to gravitational force for an electron and a proton at any distance (taking the ratio before entering numerical values shows that the distance cancels). This ratio gives some indication

of just how much larger the Coulomb force is than the gravitational force between two of the most common particles in nature.

As the example implies, gravitational force is completely negligible on a small scale, where the interactions of individual charged particles are important. On a large scale, such as between the Earth and a person, the reverse is true. Most objects are nearly electrically neutral, and so attractive and repulsive **Coulomb forces** nearly cancel. Gravitational force on a large scale dominates interactions between large objects because it is always attractive, while Coulomb forces tend to cancel.

Section Summary

- Frenchman Charles Coulomb was the first to publish the mathematical equation that describes the electrostatic force between two objects.
- Coulomb's law gives the magnitude of the force between point charges. It is **Equation:**

$$F = k \frac{|q_1 q_2|}{r^2},$$

where q_1 and q_2 are two point charges separated by a distance r , and $k \approx 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$

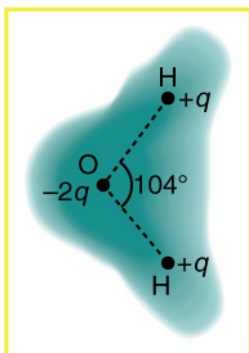
- This Coulomb force is extremely basic, since most charges are due to point-like particles. It is responsible for all electrostatic effects and underlies most macroscopic forces.
- The Coulomb force is extraordinarily strong compared with the gravitational force, another basic force—but unlike gravitational force it can cancel, since it can be either attractive or repulsive.
- The electrostatic force between two subatomic particles is far greater than the gravitational force between the same two particles.

Conceptual Questions

Exercise:

Problem:

[\[link\]](#) shows the charge distribution in a water molecule, which is called a polar molecule because it has an inherent separation of charge. Given water's polar character, explain what effect humidity has on removing excess charge from objects.



Schematic representation of the outer electron cloud of a neutral water molecule. The electrons spend more time near the oxygen than the hydrogens, giving a permanent charge separation as shown. Water is thus a *polar molecule*. It is more easily affected by electrostatic forces than molecules with uniform charge distributions.

Exercise:

Problem:

Using [\[link\]](#), explain, in terms of Coulomb's law, why a polar molecule (such as in [\[link\]](#)) is attracted by both positive and negative charges.

Exercise:

Problem:

Given the polar character of water molecules, explain how ions in the air form nucleation centers for rain droplets.

Problems & Exercises

Exercise:

Problem:

What is the repulsive force between two pith balls that are 8.00 cm apart and have equal charges of -30.0 nC ?

Exercise:

Problem:

(a) How strong is the attractive force between a glass rod with a $0.700\text{ }\mu\text{C}$ charge and a silk cloth with a $-0.600\text{ }\mu\text{C}$ charge, which are 12.0 cm apart, using the approximation that they act like point charges? (b) Discuss how the answer to this problem might be affected if the charges are distributed over some area and do not act like point charges.

Solution:

(a) 0.263 N

(b) If the charges are distributed over some area, there will be a concentration of charge along the side closest to the oppositely charged object. This effect will increase the net force.

Exercise:

Problem:

Two point charges exert a 5.00 N force on each other. What will the force become if the distance between them is increased by a factor of three?

Exercise:

Problem:

Two point charges are brought closer together, increasing the force between them by a factor of 25. By what factor was their separation decreased?

Solution:

The separation decreased by a factor of 5.

Exercise:**Problem:**

How far apart must two point charges of 75.0 nC (typical of static electricity) be to have a force of 1.00 N between them?

Exercise:**Problem:**

If two equal charges each of 1 C each are separated in air by a distance of 1 km, what is the magnitude of the force acting between them? You will see that even at a distance as large as 1 km, the repulsive force is substantial because 1 C is a very significant amount of charge.

Exercise:**Problem:**

A test charge of $+2\ \mu\text{C}$ is placed halfway between a charge of $+6\ \mu\text{C}$ and another of $+4\ \mu\text{C}$ separated by 10 cm. (a) What is the magnitude of the force on the test charge? (b) What is the direction of this force (away from or toward the $+6\ \mu\text{C}$ charge)?

Exercise:**Problem:**

Bare free charges do not remain stationary when close together. To illustrate this, calculate the acceleration of two isolated protons separated by 2.00 nm (a typical distance between gas atoms). Explicitly show how you follow the steps in the Problem-Solving Strategy for electrostatics.

Solution:

$$\begin{aligned} F &= k \frac{|q_1 q_2|}{r^2} = ma \Rightarrow a = \frac{kq^2}{mr^2} \\ &= \frac{(9.00 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2)(1.60 \times 10^{-19} \text{ m})^2}{(1.67 \times 10^{-27} \text{ kg})(2.00 \times 10^{-9} \text{ m})^2} \\ &= 3.45 \times 10^{16} \text{ m/s}^2 \end{aligned}$$

Exercise:**Problem:**

(a) By what factor must you change the distance between two point charges to change the force between them by a factor of 10? (b) Explain how the distance can either increase or decrease by this factor and still cause a factor of 10 change in the force.

Solution:

(a) 3.2

(b) If the distance increases by 3.2, then the force will decrease by a factor of 10 ; if the distance decreases by 3.2, then the force will increase by a factor of 10. Either way, the force changes by a factor of 10.

Exercise:

Problem:

Suppose you have a total charge q_{tot} that you can split in any manner. Once split, the separation distance is fixed. How do you split the charge to achieve the greatest force?

Exercise:

Problem:

(a) Common transparent tape becomes charged when pulled from a dispenser. If one piece is placed above another, the repulsive force can be great enough to support the top piece's weight. Assuming equal point charges (only an approximation), calculate the magnitude of the charge if electrostatic force is great enough to support the weight of a 10.0 mg piece of tape held 1.00 cm above another. (b) Discuss whether the magnitude of this charge is consistent with what is typical of static electricity.

Solution:

(a) $1.04 \times 10^{-9} \text{ C}$

(b) This charge is approximately 1 nC, which is consistent with the magnitude of charge typical for static electricity

Exercise:

Problem:

(a) Find the ratio of the electrostatic to gravitational force between two electrons. (b) What is this ratio for two protons? (c) Why is the ratio different for electrons and protons?

Exercise:

Problem:

At what distance is the electrostatic force between two protons equal to the weight of one proton?

Exercise:

Problem:

A certain five cent coin contains 5.00 g of nickel. What fraction of the nickel atoms' electrons, removed and placed 1.00 m above it, would support the weight of this coin? The atomic mass of nickel is 58.7, and each nickel atom contains 28 electrons and 28 protons.

Solution:

$$1.02 \times 10^{-11}$$

Exercise:**Problem:**

(a) Two point charges totaling $8.00 \mu\text{C}$ exert a repulsive force of 0.150 N on one another when separated by 0.500 m. What is the charge on each? (b) What is the charge on each if the force is attractive?

Exercise:**Problem:**

Point charges of $5.00 \mu\text{C}$ and $-3.00 \mu\text{C}$ are placed 0.250 m apart. (a) Where can a third charge be placed so that the net force on it is zero? (b) What if both charges are positive?

Solution:

- a. 0.859 m beyond negative charge on line connecting two charges
- b. 0.109 m from lesser charge on line connecting two charges

Exercise:**Problem:**

Two point charges q_1 and q_2 are 3.00 m apart, and their total charge is $20 \mu\text{C}$. (a) If the force of repulsion between them is 0.075N, what are magnitudes of the two charges? (b) If one charge attracts the other with a force of 0.525N, what are the magnitudes of the two charges? Note that you may need to solve a quadratic equation to reach your answer.

Glossary**Coulomb's law**

the mathematical equation calculating the electrostatic force vector between two charged particles

Coulomb force

another term for the electrostatic force

electrostatic force

the amount and direction of attraction or repulsion between two charged bodies

Electric Field: Concept of a Field Revisited

- Describe a force field and calculate the strength of an electric field due to a point charge.
- Calculate the force exerted on a test charge by an electric field.
- Explain the relationship between electrical force (F) on a test charge and electrical field strength (E).

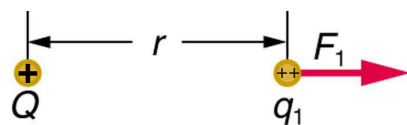
Contact forces, such as between a baseball and a bat, are explained on the small scale by the interaction of the charges in atoms and molecules in close proximity. They interact through forces that include the **Coulomb force**. Action at a distance is a force between objects that are not close enough for their atoms to “touch.” That is, they are separated by more than a few atomic diameters.

For example, a charged rubber comb attracts neutral bits of paper from a distance via the Coulomb force. It is very useful to think of an object being surrounded in space by a **force field**. The force field carries the force to another object (called a test object) some distance away.

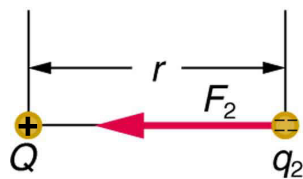
Concept of a Field

A field is a way of conceptualizing and mapping the force that surrounds any object and acts on another object at a distance without apparent physical connection. For example, the gravitational field surrounding the earth (and all other masses) represents the gravitational force that would be experienced if another mass were placed at a given point within the field.

In the same way, the Coulomb force field surrounding any charge extends throughout space. Using Coulomb’s law, $F = k|q_1q_2|/r^2$, its magnitude is given by the equation $F = k|qQ|/r^2$, for a **point charge** (a particle having a charge Q) acting on a **test charge** q at a distance r (see [\[link\]](#)). Both the magnitude and direction of the Coulomb force field depend on Q and the test charge q .



(a)



(b)

The Coulomb force field due to a positive charge Q is shown acting on two different charges. Both charges are the same distance from Q . (a) Since q_1 is positive, the force F_1 acting on it is repulsive. (b) The charge q_2 is negative and greater in magnitude than q_1 , and so the force F_2 acting on it is attractive and stronger than F_1 . The Coulomb force field is thus not unique at any point in space, because it depends on the test charges q_1 and q_2

as well as the
charge Q .

To simplify things, we would prefer to have a field that depends only on Q and not on the test charge q . The electric field is defined in such a manner that it represents only the charge creating it and is unique at every point in space. Specifically, the electric field E is defined to be the ratio of the Coulomb force to the test charge:

Equation:

$$\mathbf{E} = \frac{\mathbf{F}}{q},$$

where \mathbf{F} is the electrostatic force (or Coulomb force) exerted on a positive test charge q . It is understood that \mathbf{E} is in the same direction as \mathbf{F} . It is also assumed that q is so small that it does not alter the charge distribution creating the electric field. The units of electric field are newtons per coulomb (N/C). If the electric field is known, then the electrostatic force on any charge q is simply obtained by multiplying charge times electric field, or $\mathbf{F} = q\mathbf{E}$. Consider the electric field due to a point charge Q . According to Coulomb's law, the force it exerts on a test charge q is $F = k|qQ|/r^2$. Thus the magnitude of the electric field, E , for a point charge is

Equation:

$$E = \left| \frac{F}{q} \right| = k \left| \frac{qQ}{qr^2} \right| = k \frac{|Q|}{r^2}.$$

Since the test charge cancels, we see that

Equation:

$$E = k \frac{|Q|}{r^2}.$$

The electric field is thus seen to depend only on the charge Q and the distance r ; it is completely independent of the test charge q .

Example:

Calculating the Electric Field of a Point Charge

Calculate the strength and direction of the electric field E due to a point charge of 2.00 nC (nano-Coulombs) at a distance of 5.00 mm from the charge.

Strategy

We can find the electric field created by a point charge by using the equation $E = kQ/r^2$.

Solution

Here $Q = 2.00 \times 10^{-9} \text{ C}$ and $r = 5.00 \times 10^{-3} \text{ m}$. Entering those values into the above equation gives

Equation:

$$\begin{aligned} E &= k \frac{Q}{r^2} \\ &= (8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \times \frac{(2.00 \times 10^{-9} \text{ C})}{(5.00 \times 10^{-3} \text{ m})^2} \\ &= 7.19 \times 10^5 \text{ N/C}. \end{aligned}$$

Discussion

This **electric field strength** is the same at any point 5.00 mm away from the charge Q that creates the field. It is positive, meaning that it has a direction pointing away from the charge Q .

Example:

Calculating the Force Exerted on a Point Charge by an Electric Field

What force does the electric field found in the previous example exert on a point charge of $-0.250 \mu\text{C}$?

Strategy

Since we know the electric field strength and the charge in the field, the force on that charge can be calculated using the definition of electric field

$\mathbf{E} = \mathbf{F}/q$ rearranged to $\mathbf{F} = q\mathbf{E}$.

Solution

The magnitude of the force on a charge $q = -0.250 \mu\text{C}$ exerted by a field of strength $E = 7.20 \times 10^5 \text{ N/C}$ is thus,

Equation:

$$\begin{aligned} F &= -qE \\ &= (0.250 \times 10^{-6} \text{ C})(7.20 \times 10^5 \text{ N/C}) \\ &= 0.180 \text{ N.} \end{aligned}$$

Because q is negative, the force is directed opposite to the direction of the field.

Discussion

The force is attractive, as expected for unlike charges. (The field was created by a positive charge and here acts on a negative charge.) The charges in this example are typical of common static electricity, and the modest attractive force obtained is similar to forces experienced in static cling and similar situations.

Note:

PhET Explorations: Electric Field of Dreams

Play ball! Add charges to the Field of Dreams and see how they react to the electric field. Turn on a background electric field and adjust the direction and magnitude.

<https://archive.cnx.org/specials/ca9a78b4-06a7-11e6-b638-3bb71d1f0b42/electric-field-of-dreams/#sim-electric-field-of-dreams>

Section Summary

- The electrostatic force field surrounding a charged object extends out into space in all directions.
- The electrostatic force exerted by a point charge on a test charge at a distance r depends on the charge of both charges, as well as the

distance between the two.

- The electric field \mathbf{E} is defined to be
Equation:

$$\mathbf{E} = \frac{\mathbf{F}}{q},$$

where \mathbf{F} is the Coulomb or electrostatic force exerted on a small positive test charge q . \mathbf{E} has units of N/C.

- The magnitude of the electric field \mathbf{E} created by a point charge Q is
Equation:

$$\mathbf{E} = k \frac{|Q|}{r^2}.$$

where r is the distance from Q . The electric field \mathbf{E} is a vector and fields due to multiple charges add like vectors.

Conceptual Questions

Exercise:

Problem:

Why must the test charge q in the definition of the electric field be vanishingly small?

Exercise:

Problem:

Are the direction and magnitude of the Coulomb force unique at a given point in space? What about the electric field?

Problem Exercises

Exercise:

Problem:

What is the magnitude and direction of an electric field that exerts a 2.00×10^{-5} N upward force on a $-1.75 \mu\text{C}$ charge?

Exercise:**Problem:**

What is the magnitude and direction of the force exerted on a $3.50 \mu\text{C}$ charge by a 250 N/C electric field that points due east?

Solution:

$$8.75 \times 10^{-4} \text{ N}$$

Exercise:**Problem:**

Calculate the magnitude of the electric field 2.00 m from a point charge of 5.00 mC (such as found on the terminal of a Van de Graaff).

Exercise:**Problem:**

(a) What magnitude point charge creates a 10,000 N/C electric field at a distance of 0.250 m? (b) How large is the field at 10.0 m?

Solution:

(a) $6.94 \times 10^{-8} \text{ C}$

(b) 6.25 N/C

Exercise:

Problem:

Calculate the initial (from rest) acceleration of a proton in a $5.00 \times 10^6 \text{ N/C}$ electric field (such as created by a research Van de Graaff). Explicitly show how you follow the steps in the Problem-Solving Strategy for electrostatics.

Exercise:**Problem:**

(a) Find the magnitude and direction of an electric field that exerts a $4.80 \times 10^{-17} \text{ N}$ westward force on an electron. (b) What magnitude and direction force does this field exert on a proton?

Solution:

(a) 300 N/C (east)

(b) $4.80 \times 10^{-17} \text{ N}$ (east)

Glossary**field**

a map of the amount and direction of a force acting on other objects, extending out into space

point charge

A charged particle, designated Q , generating an electric field

test charge

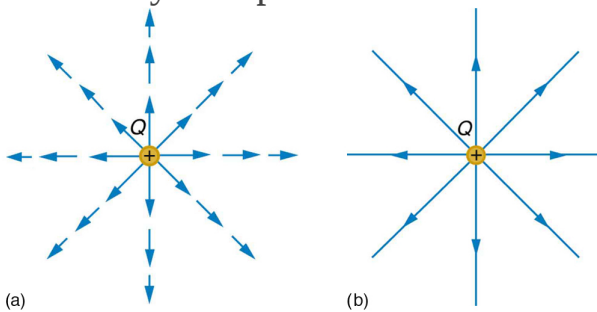
A particle (designated q) with either a positive or negative charge set down within an electric field generated by a point charge

Electric Field Lines: Multiple Charges

- Calculate the total force (magnitude and direction) exerted on a test charge from more than one charge
- Describe an electric field diagram of a positive point charge; of a negative point charge with twice the magnitude of positive charge
- Draw the electric field lines between two points of the same charge; between two points of opposite charge.

Drawings using lines to represent **electric fields** around charged objects are very useful in visualizing field strength and direction. Since the electric field has both magnitude and direction, it is a vector. Like all **vectors**, the electric field can be represented by an arrow that has length proportional to its magnitude and that points in the correct direction. (We have used arrows extensively to represent force vectors, for example.)

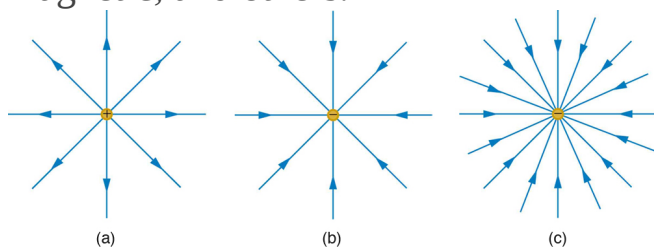
[\[link\]](#) shows two pictorial representations of the same electric field created by a positive point charge Q . [\[link\]](#) (b) shows the standard representation using continuous lines. [\[link\]](#) (a) shows numerous individual arrows with each arrow representing the force on a test charge q . Field lines are essentially a map of infinitesimal force vectors.



Two equivalent representations of the electric field due to a positive charge Q . (a) Arrows representing the electric field's magnitude and direction. (b) In the standard representation, the arrows are replaced by continuous field lines having the same direction at any point

as the electric field. The closeness of the lines is directly related to the strength of the electric field. A test charge placed anywhere will feel a force in the direction of the field line; this force will have a strength proportional to the density of the lines (being greater near the charge, for example).

Note that the electric field is defined for a positive test charge q , so that the field lines point away from a positive charge and toward a negative charge. (See [\[link\]](#).) The electric field strength is exactly proportional to the number of field lines per unit area, since the magnitude of the electric field for a point charge is $E = k|Q|/r^2$ and area is proportional to r^2 . This pictorial representation, in which field lines represent the direction and their closeness (that is, their areal density or the number of lines crossing a unit area) represents strength, is used for all fields: electrostatic, gravitational, magnetic, and others.



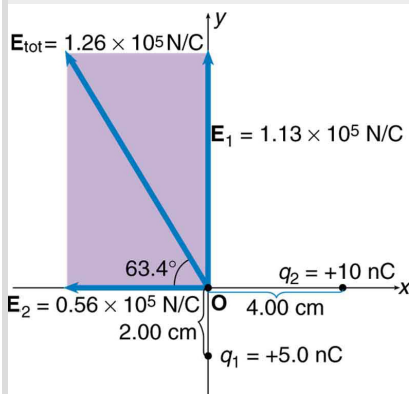
The electric field surrounding three different point charges. (a) A positive charge. (b) A negative charge of equal magnitude. (c) A larger negative charge.

In many situations, there are multiple charges. The total electric field created by multiple charges is the vector sum of the individual fields created by each charge. The following example shows how to add electric field vectors.

Example:

Adding Electric Fields

Find the magnitude and direction of the total electric field due to the two point charges, q_1 and q_2 , at the origin of the coordinate system as shown in [\[link\]](#).



The electric fields \mathbf{E}_1 and \mathbf{E}_2 at the origin O add to \mathbf{E}_{tot} .

Strategy

Since the electric field is a vector (having magnitude and direction), we add electric fields with the same vector techniques used for other types of vectors. We first must find the electric field due to each charge at the point of interest, which is the origin of the coordinate system (O) in this instance. We pretend that there is a positive test charge, q , at point O , which allows us to determine the direction of the fields \mathbf{E}_1 and \mathbf{E}_2 . Once those fields are found, the total field can be determined using **vector addition**.

Solution

The electric field strength at the origin due to q_1 is labeled E_1 and is calculated:

Equation:

$$E_1 = k \frac{q_1}{r_1^2} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2 / \text{C}^2 \right) \frac{(5.00 \times 10^{-9} \text{ C})}{(2.00 \times 10^{-2} \text{ m})^2}$$
$$E_1 = 1.124 \times 10^5 \text{ N/C}.$$

Similarly, E_2 is

Equation:

$$E_2 = k \frac{q_2}{r_2^2} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2 / \text{C}^2 \right) \frac{(10.0 \times 10^{-9} \text{ C})}{(4.00 \times 10^{-2} \text{ m})^2}$$
$$E_2 = 0.5619 \times 10^5 \text{ N/C}.$$

Four digits have been retained in this solution to illustrate that E_1 is exactly twice the magnitude of E_2 . Now arrows are drawn to represent the magnitudes and directions of \mathbf{E}_1 and \mathbf{E}_2 . (See [\[link\]](#).) The direction of the electric field is that of the force on a positive charge so both arrows point directly away from the positive charges that create them. The arrow for \mathbf{E}_1 is exactly twice the length of that for \mathbf{E}_2 . The arrows form a right triangle in this case and can be added using the Pythagorean theorem. The magnitude of the total field E_{tot} is

Equation:

$$E_{\text{tot}} = (E_1^2 + E_2^2)^{1/2}$$
$$= \{(1.124 \times 10^5 \text{ N/C})^2 + (0.5619 \times 10^5 \text{ N/C})^2\}^{1/2}$$
$$= 1.26 \times 10^5 \text{ N/C}.$$

The direction is

Equation:

$$\begin{aligned}
 \theta &= \tan^{-1} \left(\frac{E_1}{E_2} \right) \\
 &= \tan^{-1} \left(\frac{1.124 \times 10^5 \text{ N/C}}{0.5619 \times 10^5 \text{ N/C}} \right) \\
 &= 63.4^\circ,
 \end{aligned}$$

or 63.4° above the x -axis.

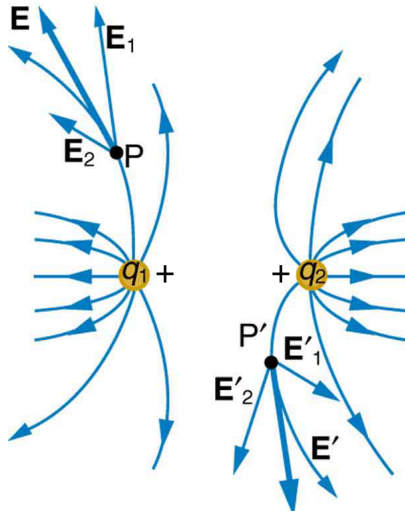
Discussion

In cases where the electric field vectors to be added are not perpendicular, vector components or graphical techniques can be used. The total electric field found in this example is the total electric field at only one point in space. To find the total electric field due to these two charges over an entire region, the same technique must be repeated for each point in the region. This impossibly lengthy task (there are an infinite number of points in space) can be avoided by calculating the total field at representative points and using some of the unifying features noted next.

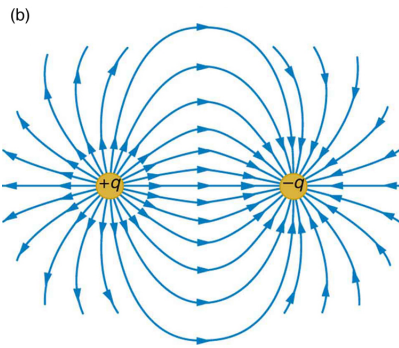
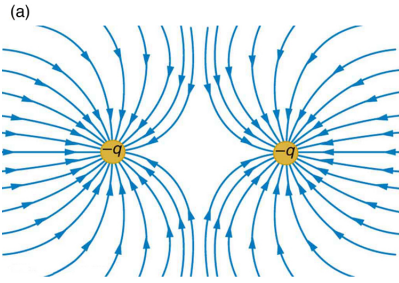
[\[link\]](#) shows how the electric field from two point charges can be drawn by finding the total field at representative points and drawing electric field lines consistent with those points. While the electric fields from multiple charges are more complex than those of single charges, some simple features are easily noticed.

For example, the field is weaker between like charges, as shown by the lines being farther apart in that region. (This is because the fields from each charge exert opposing forces on any charge placed between them.) (See [\[link\]](#) and [\[link\]](#)(a).) Furthermore, at a great distance from two like charges, the field becomes identical to the field from a single, larger charge.

[\[link\]](#)(b) shows the electric field of two unlike charges. The field is stronger between the charges. In that region, the fields from each charge are in the same direction, and so their strengths add. The field of two unlike charges is weak at large distances, because the fields of the individual charges are in opposite directions and so their strengths subtract. At very large distances, the field of two unlike charges looks like that of a smaller single charge.



Two positive point charges q_1 and q_2 produce the resultant electric field shown. The field is calculated at representative points and then smooth field lines drawn following the rules outlined in the text.



(a) Two negative charges produce the fields shown. It is very similar to the field produced by two positive charges, except that the directions are reversed. The field is clearly weaker between the charges. The individual forces on a test charge in that region are in opposite directions.

(b) Two opposite charges produce the field shown, which is stronger in the region between the charges.

We use electric field lines to visualize and analyze electric fields (the lines are a pictorial tool, not a physical entity in themselves). The properties of electric field lines for any charge distribution can be summarized as follows:

1. Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
2. The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
3. The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
4. The direction of the electric field is tangent to the field line at any point in space.
5. Field lines can never cross.

The last property means that the field is unique at any point. The field line represents the direction of the field; so if they crossed, the field would have two directions at that location (an impossibility if the field is unique).

Note:

PhET Explorations: Charges and Fields

Move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more. It's colorful, it's dynamic, it's free.

[Click here
for the
simulation](#)

•

Section Summary

- Drawings of electric field lines are useful visual tools. The properties of electric field lines for any charge distribution are that:
- Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
- The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
- The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
- The direction of the electric field is tangent to the field line at any point in space.
- Field lines can never cross.

Conceptual Questions

Exercise:

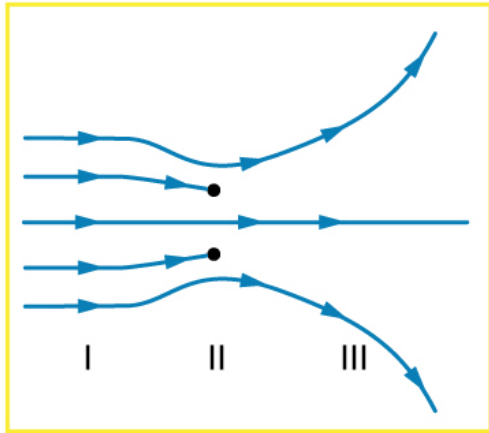
Problem:

Compare and contrast the Coulomb force field and the electric field. To do this, make a list of five properties for the Coulomb force field analogous to the five properties listed for electric field lines. Compare each item in your list of Coulomb force field properties with those of the electric field—are they the same or different? (For example, electric field lines cannot cross. Is the same true for Coulomb field lines?)

Exercise:

Problem:

[\[link\]](#) shows an electric field extending over three regions, labeled I, II, and III. Answer the following questions. (a) Are there any isolated charges? If so, in what region and what are their signs? (b) Where is the field strongest? (c) Where is it weakest? (d) Where is the field the most uniform?



Problem Exercises

Exercise:

Problem:

(a) Sketch the electric field lines near a point charge $+q$. (b) Do the same for a point charge $-3.00q$.

Exercise:

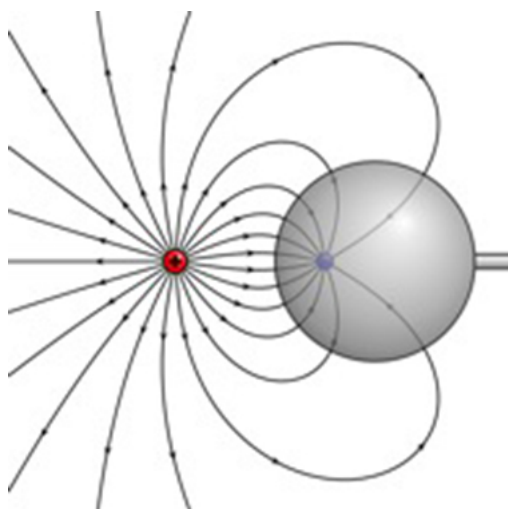
Problem:

Sketch the electric field lines a long distance from the charge distributions shown in [\[link\]](#) (a) and (b)

Exercise:

Problem:

[\[link\]](#) shows the electric field lines near two charges q_1 and q_2 . What is the ratio of their magnitudes? (b) Sketch the electric field lines a long distance from the charges shown in the figure.



The electric field near two charges.

Exercise:

Problem:

Sketch the electric field lines in the vicinity of two opposite charges, where the negative charge is three times greater in magnitude than the positive. (See [\[link\]](#) for a similar situation).

Glossary

electric field

a three-dimensional map of the electric force extended out into space from a point charge

electric field lines

a series of lines drawn from a point charge representing the magnitude and direction of force exerted by that charge

vector

a quantity with both magnitude and direction

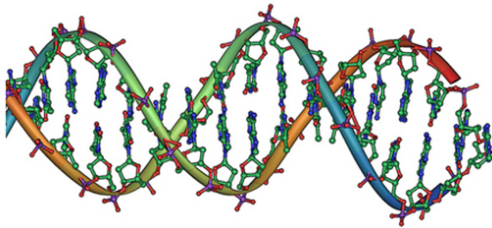
vector addition

mathematical combination of two or more vectors, including their magnitudes, directions, and positions

Electric Forces in Biology

- Describe how a water molecule is polar.
- Explain electrostatic screening by a water molecule within a living cell.

Classical electrostatics has an important role to play in modern molecular biology. Large molecules such as proteins, nucleic acids, and so on—so important to life—are usually electrically charged. DNA itself is highly charged; it is the electrostatic force that not only holds the molecule together but gives the molecule structure and strength. [\[link\]](#) is a schematic of the DNA double helix.



DNA is a highly charged molecule. The DNA double helix shows the two coiled strands each containing a row of nitrogenous bases, which “code” the genetic information needed by a living organism. The strands are connected by bonds between pairs of bases. While pairing combinations between certain bases are fixed (C-G and A-T), the sequence of nucleotides in the strand varies. (credit: Jerome Walker)

The four nucleotide bases are given the symbols A (adenine), C (cytosine), G (guanine), and T (thymine). The order of the four bases varies in each strand, but the pairing between bases is always the same. C and G are always paired and A and T are always paired, which helps to preserve the order of bases in cell division (mitosis) so as to pass on the correct genetic information. Since the Coulomb force drops with distance ($F \propto 1/r^2$), the distances between the base pairs must be small enough that the electrostatic force is sufficient to hold them together.

DNA is a highly charged molecule, with about $2q_e$ (fundamental charge) per 0.3×10^{-9} m. The distance separating the two strands that make up the DNA structure is about 1 nm, while the distance separating the individual atoms within each base is about 0.3 nm.

One might wonder why electrostatic forces do not play a larger role in biology than they do if we have so many charged molecules. The reason is that the electrostatic force is “diluted” due to **screening** between molecules. This is due to the presence of other charges in the cell.

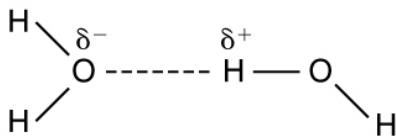
Polarity of Water Molecules

The best example of this charge screening is the water molecule, represented as H_2O . Water is a strongly **polar molecule**. Its 10 electrons (8 from the oxygen atom and 2 from the two hydrogen atoms) tend to remain closer to the oxygen nucleus than the hydrogen nuclei. This creates two centers of equal and opposite charges—what is called a **dipole**, as illustrated in [\[link\]](#). The magnitude of the dipole is called the dipole moment.

These two centers of charge will terminate some of the electric field lines coming from a free charge, as on a DNA molecule. This results in a reduction in the strength of the **Coulomb interaction**. One might say that screening makes the Coulomb force a short range force rather than long range.

Other ions of importance in biology that can reduce or screen Coulomb interactions are Na^+ , and K^+ , and Cl^- . These ions are located both inside and outside of living cells. The movement of these ions through cell membranes is crucial to the motion of nerve impulses through nerve axons.

Recent studies of electrostatics in biology seem to show that electric fields in cells can be extended over larger distances, in spite of screening, by “microtubules” within the cell. These microtubules are hollow tubes composed of proteins that guide the movement of chromosomes when cells divide, the motion of other organisms within the cell, and provide mechanisms for motion of some cells (as motors).



This schematic shows water (H_2O) as a polar molecule. Unequal sharing of electrons between the oxygen (O) and hydrogen (H) atoms leads to a net separation of positive and negative charge—forming a dipole.

The symbols δ^- and δ^+ indicate that the oxygen side of the H_2O molecule tends to be more negative, while the hydrogen ends tend

to be more positive.

This leads to an attraction of opposite charges between molecules.

Section Summary

- Many molecules in living organisms, such as DNA, carry a charge.
- An uneven distribution of the positive and negative charges within a polar molecule produces a dipole.
- The effect of a Coulomb field generated by a charged object may be reduced or blocked by other nearby charged objects.
- Biological systems contain water, and because water molecules are polar, they have a strong effect on other molecules in living systems.

Conceptual Question

Exercise:

Problem:

A cell membrane is a thin layer enveloping a cell. The thickness of the membrane is much less than the size of the cell. In a static situation the membrane has a charge distribution of $-2.5 \times 10^{-6} \text{C/m}^2$ on its inner surface and $+2.5 \times 10^{-6} \text{C/m}^2$ on its outer surface. Draw a diagram of the cell and the surrounding cell membrane. Include on this diagram the charge distribution and the corresponding electric field. Is there any electric field inside the cell? Is there any electric field outside the cell?

Glossary

dipole

a molecule's lack of symmetrical charge distribution, causing one side to be more positive and another to be more negative

polar molecule

a molecule with an asymmetrical distribution of positive and negative charge

screening

the dilution or blocking of an electrostatic force on a charged object by the presence of other charges nearby

Coulomb interaction

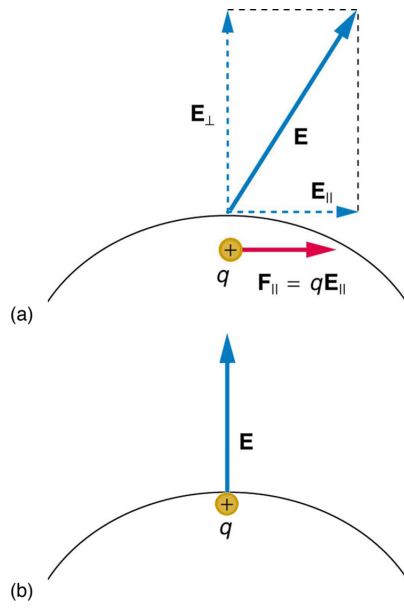
the interaction between two charged particles generated by the Coulomb forces they exert on one another

Conductors and Electric Fields in Static Equilibrium

- List the three properties of a conductor in electrostatic equilibrium.
- Explain the effect of an electric field on free charges in a conductor.
- Explain why no electric field may exist inside a conductor.
- Describe the electric field surrounding Earth.
- Explain what happens to an electric field applied to an irregular conductor.
- Describe how a lightning rod works.
- Explain how a metal car may protect passengers inside from the dangerous electric fields caused by a downed line touching the car.

Conductors contain **free charges** that move easily. When excess charge is placed on a conductor or the conductor is put into a static electric field, charges in the conductor quickly respond to reach a steady state called **electrostatic equilibrium**.

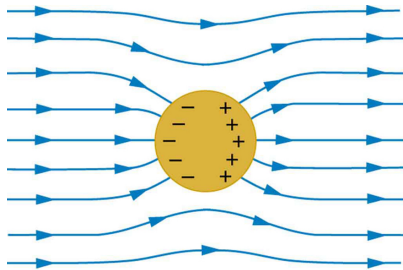
[\[link\]](#) shows the effect of an electric field on free charges in a conductor. The free charges move until the field is perpendicular to the conductor's surface. There can be no component of the field parallel to the surface in electrostatic equilibrium, since, if there were, it would produce further movement of charge. A positive free charge is shown, but free charges can be either positive or negative and are, in fact, negative in metals. The motion of a positive charge is equivalent to the motion of a negative charge in the opposite direction.



When an electric field \mathbf{E} is applied to a conductor, free charges inside the conductor move until the field is perpendicular to the surface. (a) The electric field is a vector quantity, with both parallel and perpendicular components. The parallel component (\mathbf{E}_\parallel) exerts a force (\mathbf{F}_\parallel) on the free charge q , which moves the charge until $\mathbf{F}_\parallel = 0$. (b) The resulting field is perpendicular to the surface. The free charge has

been brought to the conductor's surface, leaving electrostatic forces in equilibrium.

A conductor placed in an **electric field** will be **polarized**. [\[link\]](#) shows the result of placing a neutral conductor in an originally uniform electric field. The field becomes stronger near the conductor but entirely disappears inside it.



This illustration shows a spherical conductor in static equilibrium with an originally uniform electric field. Free charges move within the conductor, polarizing it, until the electric field lines are perpendicular to the surface. The field lines end on excess negative charge on one section of the surface and begin

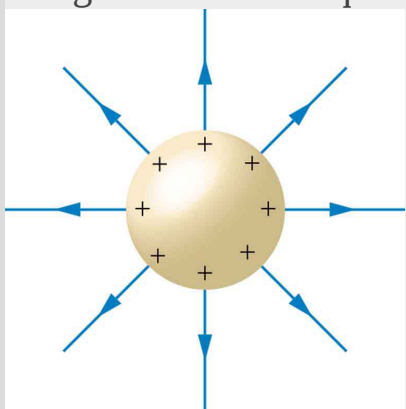
again on excess positive charge on the opposite side.

No electric field exists inside the conductor, since free charges in the conductor would continue moving in response to any field until it was neutralized.

Note:

Misconception Alert: Electric Field inside a Conductor

Excess charges placed on a spherical conductor repel and move until they are evenly distributed, as shown in [\[link\]](#). Excess charge is forced to the surface until the field inside the conductor is zero. Outside the conductor, the field is exactly the same as if the conductor were replaced by a point charge at its center equal to the excess charge.



The mutual repulsion of excess positive charges on

a spherical conductor distributes them uniformly on its surface. The resulting electric field is perpendicular to the surface and zero inside. Outside the conductor, the field is identical to that of a point charge at the center equal to the excess charge.

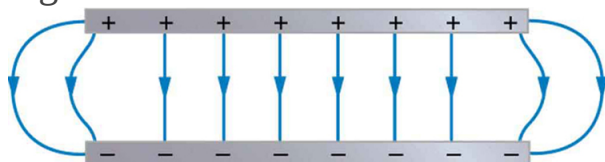
Note:

Properties of a Conductor in Electrostatic Equilibrium

1. The electric field is zero inside a conductor.
2. Just outside a conductor, the electric field lines are perpendicular to its surface, ending or beginning on charges on the surface.
3. Any excess charge resides entirely on the surface or surfaces of a conductor.

The properties of a conductor are consistent with the situations already discussed and can be used to analyze any conductor in electrostatic equilibrium. This can lead to some interesting new insights, such as described below.

How can a very uniform electric field be created? Consider a system of two metal plates with opposite charges on them, as shown in [\[link\]](#). The properties of conductors in electrostatic equilibrium indicate that the electric field between the plates will be uniform in strength and direction. Except near the edges, the excess charges distribute themselves uniformly, producing field lines that are uniformly spaced (hence uniform in strength) and perpendicular to the surfaces (hence uniform in direction, since the plates are flat). The edge effects are less important when the plates are close together.



Two metal plates with equal,
but opposite, excess charges.

The field between them is
uniform in strength and
direction except near the edges.

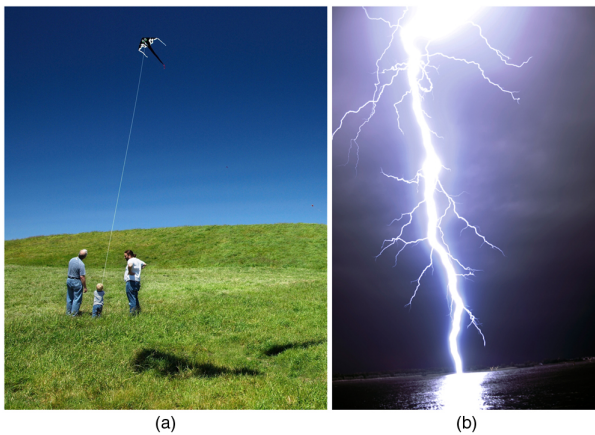
One use of such a field is to
produce uniform acceleration of
charges between the plates,
such as in the electron gun of a
TV tube.

Earth's Electric Field

A near uniform electric field of approximately 150 N/C , directed downward, surrounds Earth, with the magnitude increasing slightly as we get closer to the surface. What causes the electric field? At around 100 km above the surface of Earth we have a layer of charged particles, called the **ionosphere**. The ionosphere is responsible for a range of phenomena including the electric field surrounding Earth. In fair weather the ionosphere is positive and the Earth largely negative, maintaining the electric field ([\[link\]](#)(a)).

In storm conditions clouds form and localized electric fields can be larger and reversed in direction ([link](#)(b)). The exact charge distributions depend on the local conditions, and variations of [link](#)(b) are possible.

If the electric field is sufficiently large, the insulating properties of the surrounding material break down and it becomes conducting. For air this occurs at around 3×10^6 N/C. Air ionizes ions and electrons recombine, and we get discharge in the form of lightning sparks and corona discharge.



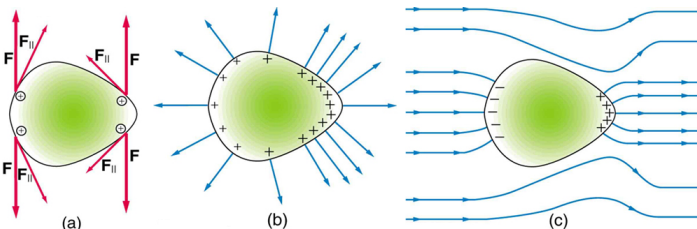
Earth's electric field. (a) Fair weather field. Earth and the ionosphere (a layer of charged particles) are both conductors. They produce a uniform electric field of about 150 N/C. (credit: D. H. Parks) (b) Storm fields. In the presence of storm clouds, the local electric fields can be larger. At very high fields, the insulating properties of the air break down and lightning can occur. (credit: Jan-Joost Verhoef)

Electric Fields on Uneven Surfaces

So far we have considered excess charges on a smooth, symmetrical conductor surface. What happens if a conductor has sharp corners or is pointed? Excess charges on a nonuniform conductor become concentrated at the sharpest points. Additionally, excess charge may move on or off the conductor at the sharpest points.

To see how and why this happens, consider the charged conductor in [\[link\]](#). The electrostatic repulsion of like charges is most effective in moving them apart on the flattest surface, and so they become least concentrated there. This is because the forces between identical pairs of charges at either end of the conductor are identical, but the components of the forces parallel to the surfaces are different. The component parallel to the surface is greatest on the flattest surface and, hence, more effective in moving the charge.

The same effect is produced on a conductor by an externally applied electric field, as seen in [\[link\]](#) (c). Since the field lines must be perpendicular to the surface, more of them are concentrated on the most curved parts.



Excess charge on a nonuniform conductor becomes most concentrated at the location of greatest curvature.

- (a) The forces between identical pairs of charges at either end of the conductor are identical, but the components of the forces parallel to the surface are different. It is \mathbf{F}_{\parallel} that moves the charges apart once they

have reached the surface. (b) \mathbf{F}_{\parallel} is smallest at the more pointed end, the charges are left closer together, producing the electric field shown. (c)

An uncharged conductor in an originally uniform electric field is polarized, with the most concentrated charge at its most pointed end.

Applications of Conductors

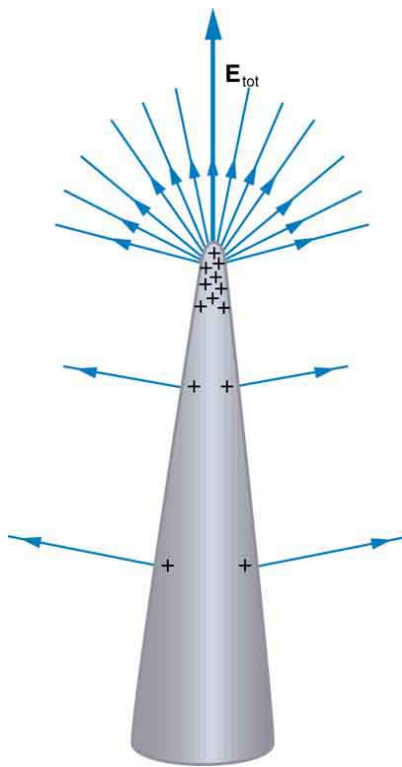
On a very sharply curved surface, such as shown in [\[link\]](#), the charges are so concentrated at the point that the resulting electric field can be great enough to remove them from the surface. This can be useful.

Lightning rods work best when they are most pointed. The large charges created in storm clouds induce an opposite charge on a building that can result in a lightning bolt hitting the building. The induced charge is bled away continually by a lightning rod, preventing the more dramatic lightning strike.

Of course, we sometimes wish to prevent the transfer of charge rather than to facilitate it. In that case, the conductor should be very smooth and have as large a radius of curvature as possible. (See [\[link\]](#).) Smooth surfaces are used on high-voltage transmission lines, for example, to avoid leakage of charge into the air.

Another device that makes use of some of these principles is a **Faraday cage**. This is a metal shield that encloses a volume. All electrical charges will reside on the outside surface of this shield, and there will be no electrical field inside. A Faraday cage is used to prohibit stray electrical fields in the environment from interfering with sensitive measurements, such as the electrical signals inside a nerve cell.

During electrical storms if you are driving a car, it is best to stay inside the car as its metal body acts as a Faraday cage with zero electrical field inside. If in the vicinity of a lightning strike, its effect is felt on the outside of the car and the inside is unaffected, provided you remain totally inside. This is also true if an active (“hot”) electrical wire was broken (in a storm or an accident) and fell on your car.



A very pointed conductor has a large charge concentration at the point. The electric field is very strong at the point and can exert a force large enough to transfer charge on or off the conductor.

Lightning rods are used to prevent the buildup of large excess charges on structures and, thus, are pointed.



(a) A lightning rod is pointed to facilitate the transfer of charge. (credit: Romaine, Wikimedia Commons) (b) This Van de Graaff generator has a smooth surface with a large radius of curvature to prevent the transfer of charge and allow a large voltage to be generated. The mutual repulsion of like charges is evident in the person's hair while touching the metal sphere. (credit: Jon 'ShakataGaNai' Davis/Wikimedia Commons).

Section Summary

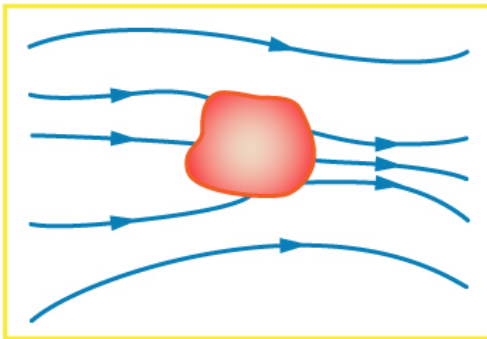
- A conductor allows free charges to move about within it.
- The electrical forces around a conductor will cause free charges to move around inside the conductor until static equilibrium is reached.
- Any excess charge will collect along the surface of a conductor.
- Conductors with sharp corners or points will collect more charge at those points.
- A lightning rod is a conductor with sharply pointed ends that collect excess charge on the building caused by an electrical storm and allow it to dissipate back into the air.
- Electrical storms result when the electrical field of Earth's surface in certain locations becomes more strongly charged, due to changes in the insulating effect of the air.
- A Faraday cage acts like a shield around an object, preventing electric charge from penetrating inside.

Conceptual Questions

Exercise:

Problem:

Is the object in [\[link\]](#) a conductor or an insulator? Justify your answer.



Exercise:

Problem:

If the electric field lines in the figure above were perpendicular to the object, would it necessarily be a conductor? Explain.

Exercise:

Problem:

The discussion of the electric field between two parallel conducting plates, in this module states that edge effects are less important if the plates are close together. What does close mean? That is, is the actual plate separation crucial, or is the ratio of plate separation to plate area crucial?

Exercise:**Problem:**

Would the self-created electric field at the end of a pointed conductor, such as a lightning rod, remove positive or negative charge from the conductor? Would the same sign charge be removed from a neutral pointed conductor by the application of a similar externally created electric field? (The answers to both questions have implications for charge transfer utilizing points.)

Exercise:**Problem:**

Why is a golfer with a metal club over her shoulder vulnerable to lightning in an open fairway? Would she be any safer under a tree?

Exercise:**Problem:**

Can the belt of a Van de Graaff accelerator be a conductor? Explain.

Exercise:**Problem:**

Are you relatively safe from lightning inside an automobile? Give two reasons.

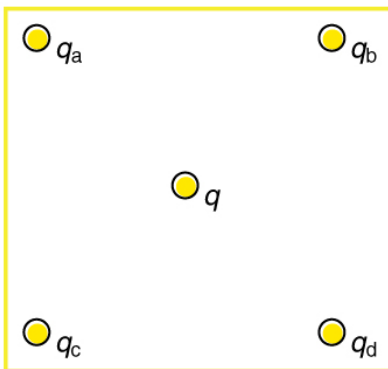
Exercise:

Problem:

Discuss pros and cons of a lightning rod being grounded versus simply being attached to a building.

Exercise:**Problem:**

Using the symmetry of the arrangement, show that the net Coulomb force on the charge q at the center of the square below ([link](#)) is zero if the charges on the four corners are exactly equal.



Four point charges q_a , q_b , q_c , and q_d lie on the corners of a square and q is located at its center.

Exercise:

Problem:

(a) Using the symmetry of the arrangement, show that the electric field at the center of the square in [\[link\]](#) is zero if the charges on the four corners are exactly equal. (b) Show that this is also true for any combination of charges in which $q_a = q_d$ and $q_b = q_c$

Exercise:**Problem:**

(a) What is the direction of the total Coulomb force on q in [\[link\]](#) if q is negative, $q_a = q_c$ and both are negative, and $q_b = q_c$ and both are positive? (b) What is the direction of the electric field at the center of the square in this situation?

Exercise:**Problem:**

Considering [\[link\]](#), suppose that $q_a = q_d$ and $q_b = q_c$. First show that q is in static equilibrium. (You may neglect the gravitational force.) Then discuss whether the equilibrium is stable or unstable, noting that this may depend on the signs of the charges and the direction of displacement of q from the center of the square.

Exercise:**Problem:**

If $q_a = 0$ in [\[link\]](#), under what conditions will there be no net Coulomb force on q ?

Exercise:**Problem:**

In regions of low humidity, one develops a special “grip” when opening car doors, or touching metal door knobs. This involves placing as much of the hand on the device as possible, not just the ends of one’s fingers. Discuss the induced charge and explain why this is done.

Exercise:**Problem:**

Tollbooth stations on roadways and bridges usually have a piece of wire stuck in the pavement before them that will touch a car as it approaches. Why is this done?

Exercise:**Problem:**

Suppose a woman carries an excess charge. To maintain her charged status can she be standing on ground wearing just any pair of shoes? How would you discharge her? What are the consequences if she simply walks away?

Problems & Exercises**Exercise:****Problem:**

Sketch the electric field lines in the vicinity of the conductor in [\[link\]](#) given the field was originally uniform and parallel to the object's long axis. Is the resulting field small near the long side of the object?

**Exercise:****Problem:**

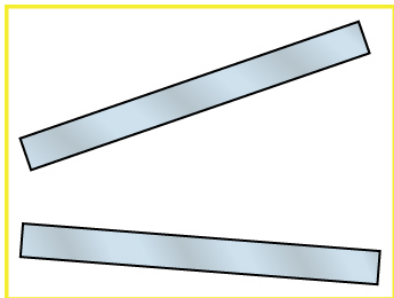
Sketch the electric field lines in the vicinity of the conductor in [\[link\]](#) given the field was originally uniform and parallel to the object's long axis. Is the resulting field small near the long side of the object?



Exercise:

Problem:

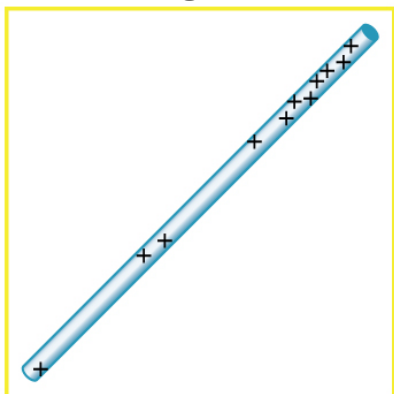
Sketch the electric field between the two conducting plates shown in [\[link\]](#), given the top plate is positive and an equal amount of negative charge is on the bottom plate. Be certain to indicate the distribution of charge on the plates.



Exercise:

Problem:

Sketch the electric field lines in the vicinity of the charged insulator in [\[link\]](#) noting its nonuniform charge distribution.



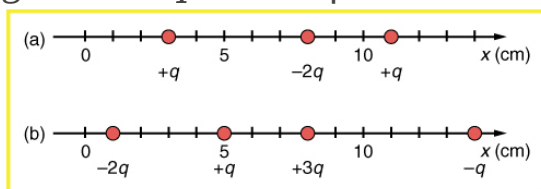
A charged
insulating rod such
as might be used in

a classroom
demonstration.

Exercise:

Problem:

What is the force on the charge located at $x = 8.00$ cm in [\[link\]](#)(a) given that $q = 1.00 \mu\text{C}$?



(a) Point charges located at 3.00, 8.00, and 11.0 cm along the x -axis. (b) Point charges located at 1.00, 5.00, 8.00, and 14.0 cm along the x -axis.

Exercise:

Problem:

(a) Find the total electric field at $x = 1.00$ cm in [\[link\]](#)(b) given that $q = 5.00 \text{ nC}$. (b) Find the total electric field at $x = 11.00$ cm in [\[link\]](#)(b). (c) If the charges are allowed to move and eventually be brought to rest by friction, what will the final charge configuration be? (That is, will there be a single charge, double charge, etc., and what will its value(s) be?)

Solution:

(a) $E_{x=1.00 \text{ cm}} = -\infty$

(b) $2.12 \times 10^5 \text{ N/C}$

(c) one charge of $+q$

Exercise:

Problem:

(a) Find the electric field at $x = 5.00 \text{ cm}$ in [\[link\]](#)(a), given that $q = 1.00 \mu\text{C}$. (b) At what position between 3.00 and 8.00 cm is the total electric field the same as that for $-2q$ alone? (c) Can the electric field be zero anywhere between 0.00 and 8.00 cm ? (d) At very large positive or negative values of x , the electric field approaches zero in both (a) and (b). In which does it most rapidly approach zero and why? (e) At what position to the right of 11.0 cm is the total electric field zero, other than at infinity? (Hint: A graphing calculator can yield considerable insight in this problem.)

Exercise:

Problem:

(a) Find the total Coulomb force on a charge of 2.00 nC located at $x = 4.00 \text{ cm}$ in [\[link\]](#) (b), given that $q = 1.00 \mu\text{C}$. (b) Find the x -position at which the electric field is zero in [\[link\]](#) (b).

Solution:

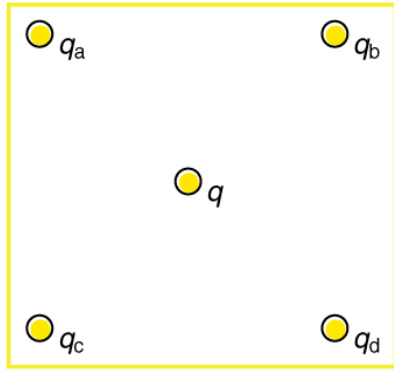
(a) 0.252 N to the left

(b) $x = 6.07 \text{ cm}$

Exercise:

Problem:

Using the symmetry of the arrangement, determine the direction of the force on q in the figure below, given that $q_a = q_b = +7.50 \mu\text{C}$ and $q_c = q_d = -7.50 \mu\text{C}$. (b) Calculate the magnitude of the force on the charge q , given that the square is 10.0 cm on a side and $q = 2.00 \mu\text{C}$.



Exercise:

Problem:

(a) Using the symmetry of the arrangement, determine the direction of the electric field at the center of the square in [\[link\]](#), given that $q_a = q_b = -1.00 \mu\text{C}$ and $q_c = q_d = +1.00 \mu\text{C}$. (b) Calculate the magnitude of the electric field at the location of q , given that the square is 5.00 cm on a side.

Solution:

(a) The electric field at the center of the square will be straight up, since q_a and q_b are positive and q_c and q_d are negative and all have the same magnitude.

(b) $2.04 \times 10^7 \text{ N/C}$ (upward)

Exercise:

Problem:

Find the electric field at the location of q_a in [\[link\]](#) given that $q_b = q_c = q_d = +2.00 \text{ nC}$, $q = -1.00 \text{ nC}$, and the square is 20.0 cm on a side.

Exercise:

Problem:

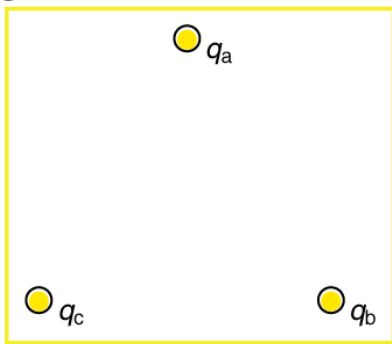
Find the total Coulomb force on the charge q in [\[link\]](#), given that $q = 1.00 \mu\text{C}$, $q_a = 2.00 \mu\text{C}$, $q_b = -3.00 \mu\text{C}$, $q_c = -4.00 \mu\text{C}$, and $q_d = +1.00 \mu\text{C}$. The square is 50.0 cm on a side.

Solution:

0.102 N, in the $-y$ direction

Exercise:**Problem:**

(a) Find the electric field at the location of q_a in [\[link\]](#), given that $q_b = +10.00 \mu\text{C}$ and $q_c = -5.00 \mu\text{C}$. (b) What is the force on q_a , given that $q_a = +1.50 \text{ nC}$?



Point charges
located at the
corners of an
equilateral triangle
25.0 cm on a side.

Exercise:

Problem:

(a) Find the electric field at the center of the triangular configuration of charges in [\[link\]](#), given that $q_a = +2.50 \text{ nC}$, $q_b = -8.00 \text{ nC}$, and $q_c = +1.50 \text{ nC}$. (b) Is there any combination of charges, other than $q_a = q_b = q_c$, that will produce a zero strength electric field at the center of the triangular configuration?

Solution:

(a) $\vec{E} = 4.36 \times 10^3 \text{ N/C}$, 35.0° , below the horizontal.

(b) No

Glossary

conductor

an object with properties that allow charges to move about freely within it

free charge

an electrical charge (either positive or negative) which can move about separately from its base molecule

electrostatic equilibrium

an electrostatically balanced state in which all free electrical charges have stopped moving about

polarized

a state in which the positive and negative charges within an object have collected in separate locations

ionosphere

a layer of charged particles located around 100 km above the surface of Earth, which is responsible for a range of phenomena including the electric field surrounding Earth

Faraday cage

a metal shield which prevents electric charge from penetrating its surface

Applications of Electrostatics

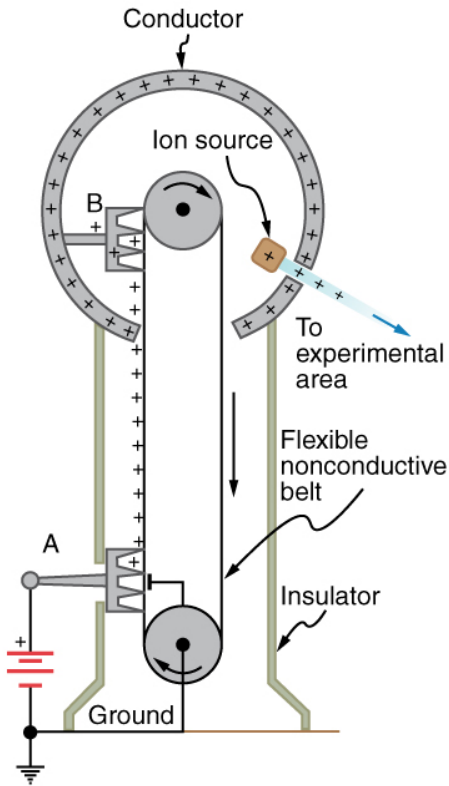
- Name several real-world applications of the study of electrostatics.

The study of **electrostatics** has proven useful in many areas. This module covers just a few of the many applications of electrostatics.

The Van de Graaff Generator

Van de Graaff generators (or Van de Graaffs) are not only spectacular devices used to demonstrate high voltage due to static electricity—they are also used for serious research. The first was built by Robert Van de Graaff in 1931 (based on original suggestions by Lord Kelvin) for use in nuclear physics research. [\[link\]](#) shows a schematic of a large research version. Van de Graaffs utilize both smooth and pointed surfaces, and conductors and insulators to generate large static charges and, hence, large voltages.

A very large excess charge can be deposited on the sphere, because it moves quickly to the outer surface. Practical limits arise because the large electric fields polarize and eventually ionize surrounding materials, creating free charges that neutralize excess charge or allow it to escape. Nevertheless, voltages of 15 million volts are well within practical limits.



Schematic of Van de Graaff generator. A battery (A) supplies excess positive charge to a pointed conductor, the points of which spray the charge onto a moving insulating belt near the bottom. The pointed conductor (B) on top in the large sphere picks up the charge. (The induced electric field at the points is so large that it removes the charge from the belt.) This can be done because the charge does not

remain inside the conducting sphere but moves to its outside surface. An ion source inside the sphere produces positive ions, which are accelerated away from the positive sphere to high velocities.

Note:

Take-Home Experiment: Electrostatics and Humidity

Rub a comb through your hair and use it to lift pieces of paper. It may help to tear the pieces of paper rather than cut them neatly. Repeat the exercise in your bathroom after you have had a long shower and the air in the bathroom is moist. Is it easier to get electrostatic effects in dry or moist air? Why would torn paper be more attractive to the comb than cut paper? Explain your observations.

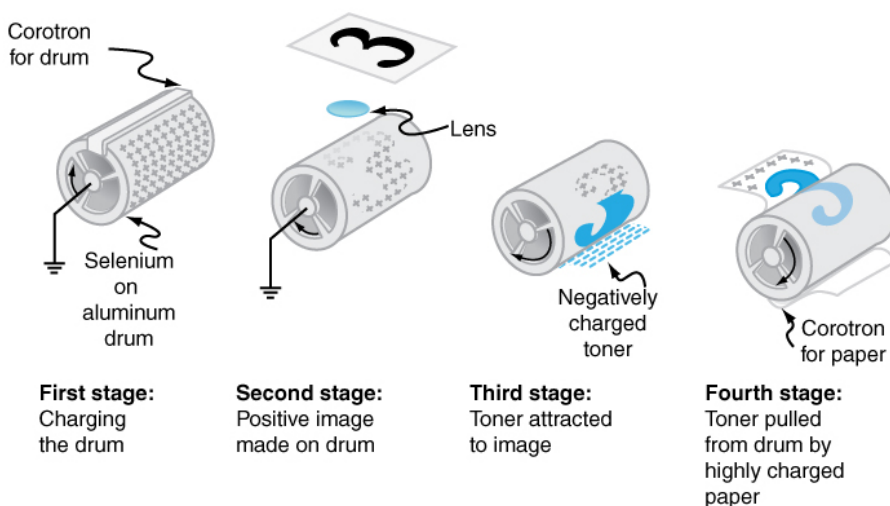
Xerography

Most copy machines use an electrostatic process called **xerography**—a word coined from the Greek words *xeros* for dry and *graphos* for writing. The heart of the process is shown in simplified form in [\[link\]](#).

A selenium-coated aluminum drum is sprayed with positive charge from points on a device called a corotron. Selenium is a substance with an interesting property—it is a **photoconductor**. That is, selenium is an insulator when in the dark and a conductor when exposed to light.

In the first stage of the xerography process, the conducting aluminum drum is **grounded** so that a negative charge is induced under the thin layer of uniformly positively charged selenium. In the second stage, the surface of the drum is exposed to the image of whatever is to be copied. Where the image is light, the selenium becomes conducting, and the positive charge is neutralized. In dark areas, the positive charge remains, and so the image has been transferred to the drum.

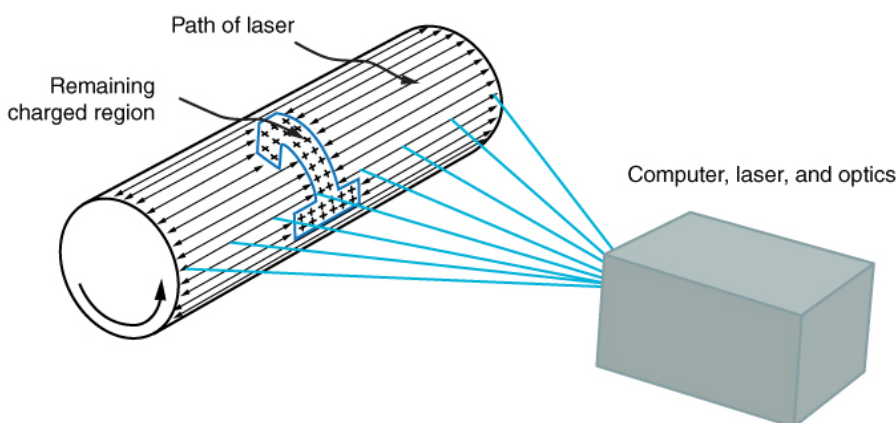
The third stage takes a dry black powder, called toner, and sprays it with a negative charge so that it will be attracted to the positive regions of the drum. Next, a blank piece of paper is given a greater positive charge than on the drum so that it will pull the toner from the drum. Finally, the paper and electrostatically held toner are passed through heated pressure rollers, which melt and permanently adhere the toner within the fibers of the paper.



Xerography is a dry copying process based on electrostatics. The major steps in the process are the charging of the photoconducting drum, transfer of an image creating a positive charge duplicate, attraction of toner to the charged parts of the drum, and transfer of toner to the paper. Not shown are heat treatment of the paper and cleansing of the drum for the next copy.

Laser Printers

Laser printers use the xerographic process to make high-quality images on paper, employing a laser to produce an image on the photoconducting drum as shown in [\[link\]](#). In its most common application, the laser printer receives output from a computer, and it can achieve high-quality output because of the precision with which laser light can be controlled. Many laser printers do significant information processing, such as making sophisticated letters or fonts, and may contain a computer more powerful than the one giving them the raw data to be printed.

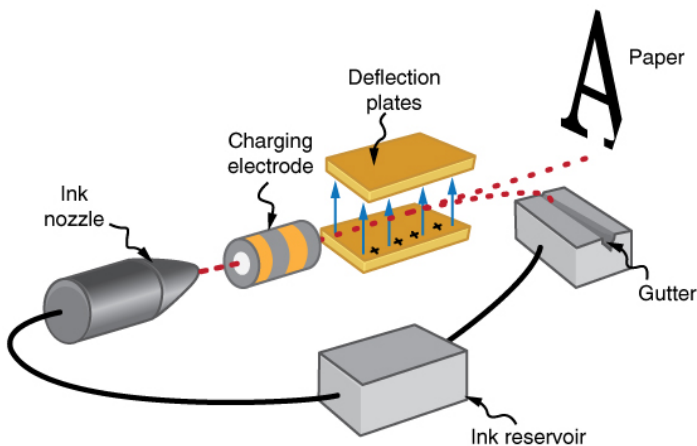


In a laser printer, a laser beam is scanned across a photoconducting drum, leaving a positive charge image. The other steps for charging the drum and transferring the image to paper are the same as in xerography. Laser light can be very precisely controlled, enabling laser printers to produce high-quality images.

Ink Jet Printers and Electrostatic Painting

The **ink jet printer**, commonly used to print computer-generated text and graphics, also employs electrostatics. A nozzle makes a fine spray of tiny ink droplets, which are then given an electrostatic charge. (See [\[link\]](#).)

Once charged, the droplets can be directed, using pairs of charged plates, with great precision to form letters and images on paper. Ink jet printers can produce color images by using a black jet and three other jets with primary colors, usually cyan, magenta, and yellow, much as a color television produces color. (This is more difficult with xerography, requiring multiple drums and toners.)



The nozzle of an ink-jet printer produces small ink droplets, which are sprayed with electrostatic charge. Various computer-driven devices are then used to direct the droplets to the correct positions on a page.

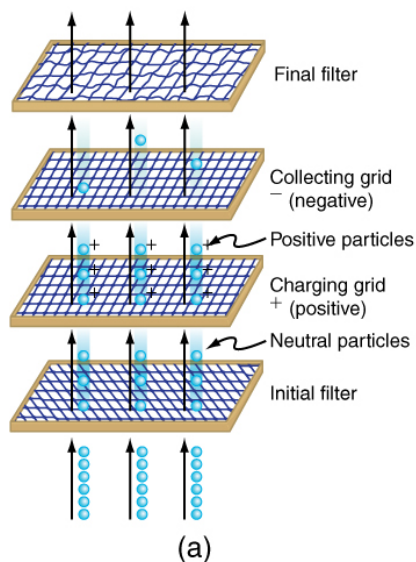
Electrostatic painting employs electrostatic charge to spray paint onto odd-shaped surfaces. Mutual repulsion of like charges causes the paint to fly away from its source. Surface tension forms drops, which are then attracted by unlike charges to the surface to be painted. Electrostatic painting can reach those hard-to-get at places, applying an even coat in a controlled

manner. If the object is a conductor, the electric field is perpendicular to the surface, tending to bring the drops in perpendicularly. Corners and points on conductors will receive extra paint. Felt can similarly be applied.

Smoke Precipitators and Electrostatic Air Cleaning

Another important application of electrostatics is found in air cleaners, both large and small. The electrostatic part of the process places excess (usually positive) charge on smoke, dust, pollen, and other particles in the air and then passes the air through an oppositely charged grid that attracts and retains the charged particles. (See [\[link\]](#).)

Large **electrostatic precipitators** are used industrially to remove over 99% of the particles from stack gas emissions associated with the burning of coal and oil. Home precipitators, often in conjunction with the home heating and air conditioning system, are very effective in removing polluting particles, irritants, and allergens.



(a) Schematic of an electrostatic precipitator. Air is passed through grids of opposite charge. The first grid charges airborne particles, while the second attracts and collects them. (b) The dramatic effect of

electrostatic precipitators is seen by the absence of smoke from this power plant. (credit: Cmdalgleish, Wikimedia Commons)

Note:

Problem-Solving Strategies for Electrostatics

1. Examine the situation to determine if static electricity is involved. This may concern separated stationary charges, the forces among them, and the electric fields they create.
2. Identify the system of interest. This includes noting the number, locations, and types of charges involved.
3. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Determine whether the Coulomb force is to be considered directly—if so, it may be useful to draw a free-body diagram, using electric field lines.
4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is important to distinguish the Coulomb force F from the electric field E , for example.
5. Solve the appropriate equation for the quantity to be determined (the unknown) or draw the field lines as requested.
6. Examine the answer to see if it is reasonable: Does it make sense? Are units correct and the numbers involved reasonable?

Integrated Concepts

The Integrated Concepts exercises for this module involve concepts such as electric charges, electric fields, and several other topics. Physics is most interesting when applied to general situations involving more than a narrow set of physical principles. The electric field exerts force on charges, for example, and hence the relevance of [Dynamics: Force and Newton's Laws of Motion](#). The following topics are involved in some or all of the problems labeled “Integrated Concepts”:

- [Kinematics](#)
- [Two-Dimensional Kinematics](#)
- [Dynamics: Force and Newton's Laws of Motion](#)
- [Uniform Circular Motion and Gravitation](#)
- [Statics and Torque](#)
- [Fluid Statics](#)

The following worked example illustrates how this strategy is applied to an Integrated Concept problem:

Example:

Acceleration of a Charged Drop of Gasoline

If steps are not taken to ground a gasoline pump, static electricity can be placed on gasoline when filling your car's tank. Suppose a tiny drop of gasoline has a mass of 4.00×10^{-15} kg and is given a positive charge of 3.20×10^{-19} C. (a) Find the weight of the drop. (b) Calculate the electric force on the drop if there is an upward electric field of strength 3.00×10^5 N/C due to other static electricity in the vicinity. (c) Calculate the drop's acceleration.

Strategy

To solve an integrated concept problem, we must first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example asks for weight. This is a topic of dynamics and is defined in [Dynamics: Force and Newton's Laws of Motion](#). Part (b) deals with electric force on a charge, a topic of [Electric Charge and Electric Field](#). Part (c) asks for acceleration, knowing forces and mass. These are part of Newton's laws, also found in [Dynamics: Force and Newton's Laws of Motion](#).

The following solutions to each part of the example illustrate how the specific problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so on.

Solution for (a)

Weight is mass times the acceleration due to gravity, as first expressed in

Equation:

$$w = mg.$$

Entering the given mass and the average acceleration due to gravity yields

Equation:

$$w = (4.00 \times 10^{-15} \text{ kg})(9.80 \text{ m/s}^2) = 3.92 \times 10^{-14} \text{ N}.$$

Discussion for (a)

This is a small weight, consistent with the small mass of the drop.

Solution for (b)

The force an electric field exerts on a charge is given by rearranging the following equation:

Equation:

$$F = qE.$$

Here we are given the charge ($3.20 \times 10^{-19} \text{ C}$ is twice the fundamental unit of charge) and the electric field strength, and so the electric force is found to be

Equation:

$$F = (3.20 \times 10^{-19} \text{ C})(3.00 \times 10^5 \text{ N/C}) = 9.60 \times 10^{-14} \text{ N}.$$

Discussion for (b)

While this is a small force, it is greater than the weight of the drop.

Solution for (c)

The acceleration can be found using Newton's second law, provided we can identify all of the external forces acting on the drop. We assume only the drop's weight and the electric force are significant. Since the drop has a positive charge and the electric field is given to be upward, the electric force is upward. We thus have a one-dimensional (vertical direction) problem, and we can state Newton's second law as

Equation:

$$a = \frac{F_{\text{net}}}{m}.$$

where $F_{\text{net}} = F - w$. Entering this and the known values into the expression for Newton's second law yields

Equation:

$$\begin{aligned} a &= \frac{F-w}{m} \\ &= \frac{9.60 \times 10^{-14} \text{ N} - 3.92 \times 10^{-14} \text{ N}}{4.00 \times 10^{-15} \text{ kg}} \\ &= 14.2 \text{ m/s}^2. \end{aligned}$$

Discussion for (c)

This is an upward acceleration great enough to carry the drop to places where you might not wish to have gasoline.

This worked example illustrates how to apply problem-solving strategies to situations that include topics in different chapters. The first step is to identify the physical principles involved in the problem. The second step is to solve for the unknown using familiar problem-solving strategies. These are found throughout the text, and many worked examples show how to use them for single topics. In this integrated concepts example, you can see how to apply them across several topics. You will find these techniques useful in applications of physics outside a physics course, such as in your profession, in other science disciplines, and in everyday life. The following problems will build your skills in the broad application of physical principles.

Note:

Unreasonable Results

The Unreasonable Results exercises for this module have results that are unreasonable because some premise is unreasonable or because certain of the premises are inconsistent with one another. Physical principles applied correctly then produce unreasonable results. The purpose of these problems is to give practice in assessing whether nature is being accurately described, and if it is not to trace the source of difficulty.

Note:**Problem-Solving Strategy**

To determine if an answer is reasonable, and to determine the cause if it is not, do the following.

1. Solve the problem using strategies as outlined above. Use the format followed in the worked examples in the text to solve the problem as usual.
2. Check to see if the answer is reasonable. Is it too large or too small, or does it have the wrong sign, improper units, and so on?
3. If the answer is unreasonable, look for what specifically could cause the identified difficulty. Usually, the manner in which the answer is unreasonable is an indication of the difficulty. For example, an extremely large Coulomb force could be due to the assumption of an excessively large separated charge.

Section Summary

- Electrostatics is the study of electric fields in static equilibrium.
- In addition to research using equipment such as a Van de Graaff generator, many practical applications of electrostatics exist, including photocopiers, laser printers, ink-jet printers and electrostatic air filters.

Problems & Exercises**Exercise:**

Problem:

(a) What is the electric field 5.00 m from the center of the terminal of a Van de Graaff with a 3.00 mC charge, noting that the field is equivalent to that of a point charge at the center of the terminal? (b) At this distance, what force does the field exert on a $2.00\ \mu\text{C}$ charge on the Van de Graaff's belt?

Exercise:**Problem:**

(a) What is the direction and magnitude of an electric field that supports the weight of a free electron near the surface of Earth? (b) Discuss what the small value for this field implies regarding the relative strength of the gravitational and electrostatic forces.

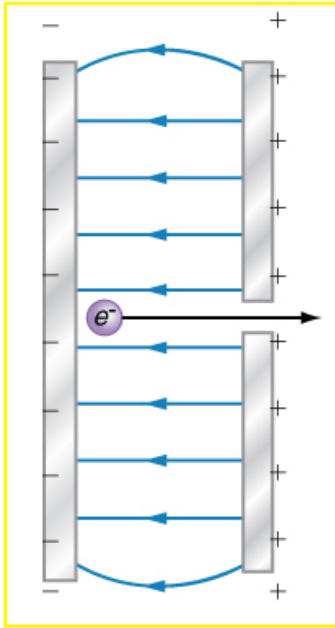
Solution:

(a) $5.58 \times 10^{-11}\ \text{N/C}$

(b) the coulomb force is extraordinarily stronger than gravity

Exercise:**Problem:**

A simple and common technique for accelerating electrons is shown in [\[link\]](#), where there is a uniform electric field between two plates. Electrons are released, usually from a hot filament, near the negative plate, and there is a small hole in the positive plate that allows the electrons to continue moving. (a) Calculate the acceleration of the electron if the field strength is $2.50 \times 10^4\ \text{N/C}$. (b) Explain why the electron will not be pulled back to the positive plate once it moves through the hole.



Parallel
conducting
plates with
opposite charges
on them create a
relatively
uniform electric
field used to
accelerate
electrons to the
right. Those that
go through the
hole can be used
to make a TV or
computer screen
glow or to
produce X-rays.

Exercise:

Problem:

Earth has a net charge that produces an electric field of approximately 150 N/C downward at its surface. (a) What is the magnitude and sign of the excess charge, noting the electric field of a conducting sphere is equivalent to a point charge at its center? (b) What acceleration will the field produce on a free electron near Earth's surface? (c) What mass object with a single extra electron will have its weight supported by this field?

Solution:

(a) $-6.76 \times 10^5 \text{ C}$

(b) $2.63 \times 10^{13} \text{ m/s}^2$ (upward)

(c) $2.45 \times 10^{-18} \text{ kg}$

Exercise:**Problem:**

Point charges of $25.0 \mu\text{C}$ and $45.0 \mu\text{C}$ are placed 0.500 m apart. (a) At what point along the line between them is the electric field zero? (b) What is the electric field halfway between them?

Exercise:**Problem:**

What can you say about two charges q_1 and q_2 , if the electric field one-fourth of the way from q_1 to q_2 is zero?

Solution:

The charge q_2 is 9 times greater than q_1 .

Exercise:**Problem: Integrated Concepts**

Calculate the angular velocity ω of an electron orbiting a proton in the hydrogen atom, given the radius of the orbit is 0.530×10^{-10} m. You may assume that the proton is stationary and the centripetal force is supplied by Coulomb attraction.

Exercise:

Problem: Integrated Concepts

An electron has an initial velocity of 5.00×10^6 m/s in a uniform 2.00×10^5 N/C strength electric field. The field accelerates the electron in the direction opposite to its initial velocity. (a) What is the direction of the electric field? (b) How far does the electron travel before coming to rest? (c) How long does it take the electron to come to rest? (d) What is the electron's velocity when it returns to its starting point?

Exercise:

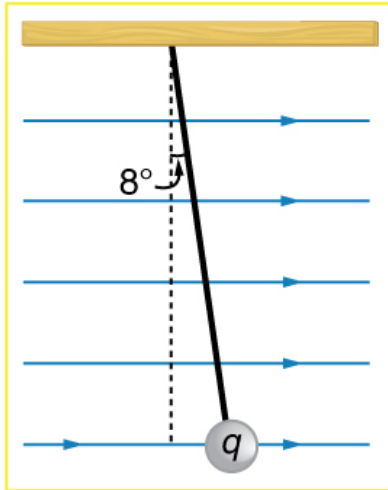
Problem: Integrated Concepts

The practical limit to an electric field in air is about 3.00×10^6 N/C. Above this strength, sparking takes place because air begins to ionize and charges flow, reducing the field. (a) Calculate the distance a free proton must travel in this field to reach 3.00% of the speed of light, starting from rest. (b) Is this practical in air, or must it occur in a vacuum?

Exercise:

Problem: Integrated Concepts

A 5.00 g charged insulating ball hangs on a 30.0 cm long string in a uniform horizontal electric field as shown in [\[link\]](#). Given the charge on the ball is $1.00 \mu\text{C}$, find the strength of the field.

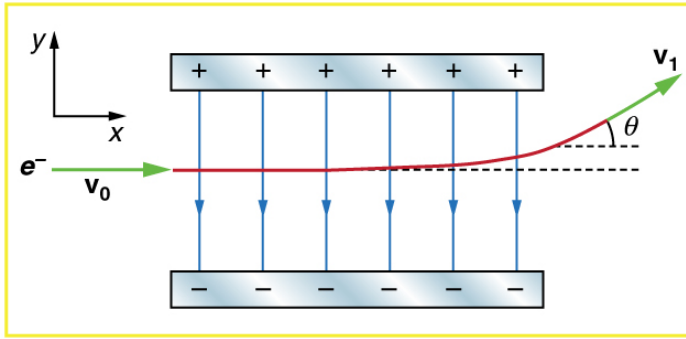


A horizontal electric field causes the charged ball to hang at an angle of 8.00° .

Exercise:

Problem: Integrated Concepts

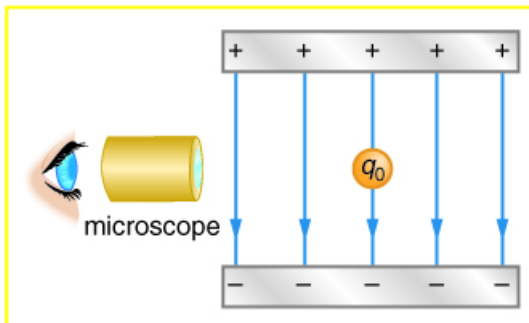
[\[link\]](#) shows an electron passing between two charged metal plates that create an 100 N/C vertical electric field perpendicular to the electron's original horizontal velocity. (These can be used to change the electron's direction, such as in an oscilloscope.) The initial speed of the electron is $3.00 \times 10^6 \text{ m/s}$, and the horizontal distance it travels in the uniform field is 4.00 cm . (a) What is its vertical deflection? (b) What is the vertical component of its final velocity? (c) At what angle does it exit? Neglect any edge effects.



Exercise:

Problem: Integrated Concepts

The classic Millikan oil drop experiment was the first to obtain an accurate measurement of the charge on an electron. In it, oil drops were suspended against the gravitational force by a vertical electric field. (See [\[link\]](#).) Given the oil drop to be $1.00 \mu\text{m}$ in radius and have a density of 920 kg/m^3 : (a) Find the weight of the drop. (b) If the drop has a single excess electron, find the electric field strength needed to balance its weight.



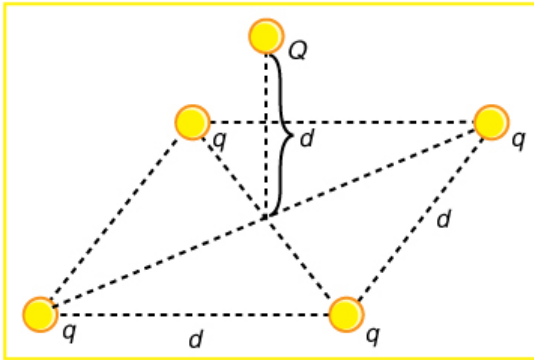
In the Millikan oil drop experiment, small drops can be suspended in an electric field by the force exerted on a single excess electron. Classically, this experiment was used to determine the electron charge q_e by

measuring the electric field
and mass of the drop.

Exercise:

Problem: Integrated Concepts

(a) In [\[link\]](#), four equal charges q lie on the corners of a square. A fifth charge Q is on a mass m directly above the center of the square, at a height equal to the length d of one side of the square. Determine the magnitude of q in terms of Q , m , and d , if the Coulomb force is to equal the weight of m . (b) Is this equilibrium stable or unstable? Discuss.



Four equal charges on the
corners of a horizontal
square support the weight of
a fifth charge located
directly above the center of
the square.

Exercise:

Problem: Unreasonable Results

(a) Calculate the electric field strength near a 10.0 cm diameter
conducting sphere that has 1.00 C of excess charge on it. (b) What is

unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

(a) Two 0.500 g raindrops in a thunderhead are 1.00 cm apart when they each acquire 1.00 mC charges. Find their acceleration. (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Exercise:

Problem: Unreasonable Results

A wrecking yard inventor wants to pick up cars by charging a 0.400 m diameter ball and inducing an equal and opposite charge on the car. If a car has a 1000 kg mass and the ball is to be able to lift it from a distance of 1.00 m: (a) What minimum charge must be used? (b) What is the electric field near the surface of the ball? (c) Why are these results unreasonable? (d) Which premise or assumption is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider two insulating balls with evenly distributed equal and opposite charges on their surfaces, held with a certain distance between the centers of the balls. Construct a problem in which you calculate the electric field (magnitude and direction) due to the balls at various points along a line running through the centers of the balls and extending to infinity on either side. Choose interesting points and comment on the meaning of the field at those points. For example, at what points might the field be just that due to one ball and where does the field become negligibly small? Among the things to be considered are the magnitudes of the charges and the distance between the centers of the balls. Your instructor may wish for you to consider the electric

field off axis or for a more complex array of charges, such as those in a water molecule.

Exercise:

Problem: Construct Your Own Problem

Consider identical spherical conducting space ships in deep space where gravitational fields from other bodies are negligible compared to the gravitational attraction between the ships. Construct a problem in which you place identical excess charges on the space ships to exactly counter their gravitational attraction. Calculate the amount of excess charge needed. Examine whether that charge depends on the distance between the centers of the ships, the masses of the ships, or any other factors. Discuss whether this would be an easy, difficult, or even impossible thing to do in practice.

Glossary

Van de Graaff generator

a machine that produces a large amount of excess charge, used for experiments with high voltage

electrostatics

the study of electric forces that are static or slow-moving

photoconductor

a substance that is an insulator until it is exposed to light, when it becomes a conductor

xerography

a dry copying process based on electrostatics

grounded

connected to the ground with a conductor, so that charge flows freely to and from the Earth to the grounded object

laser printer

uses a laser to create a photoconductive image on a drum, which attracts dry ink particles that are then rolled onto a sheet of paper to print a high-quality copy of the image

ink-jet printer

small ink droplets sprayed with an electric charge are controlled by electrostatic plates to create images on paper

electrostatic precipitators

filters that apply charges to particles in the air, then attract those charges to a filter, removing them from the airstream

Introduction to Electric Potential and Electric Energy

class="introduction"

Automated
external
defibrillato
r unit
(AED)
(credit:
U.S.
Defense
Department
photo/Tech.
Sgt.
Suzanne
M. Day)



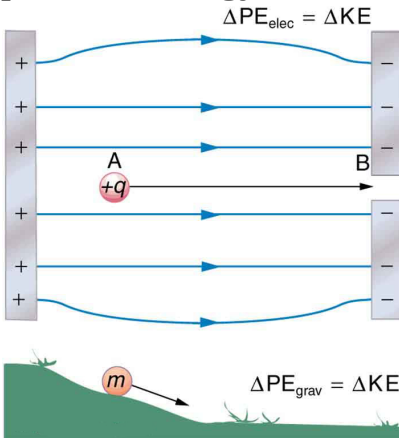
In [Electric Charge and Electric Field](#), we just scratched the surface (or at least rubbed it) of electrical phenomena. Two of the most familiar aspects of

electricity are its energy and *voltage*. We know, for example, that great amounts of electrical energy can be stored in batteries, are transmitted cross-country through power lines, and may jump from clouds to explode the sap of trees. In a similar manner, at molecular levels, *ions* cross cell membranes and transfer information. We also know about voltages associated with electricity. Batteries are typically a few volts, the outlets in your home produce 120 volts, and power lines can be as high as hundreds of thousands of volts. But energy and voltage are not the same thing. A motorcycle battery, for example, is small and would not be very successful in replacing the much larger car battery, yet each has the same voltage. In this chapter, we shall examine the relationship between voltage and electrical energy and begin to explore some of the many applications of electricity.

Electric Potential Energy: Potential Difference

- Define electric potential and electric potential energy.
- Describe the relationship between potential difference and electrical potential energy.
- Explain electron volt and its usage in submicroscopic process.
- Determine electric potential energy given potential difference and amount of charge.

When a free positive charge q is accelerated by an electric field, such as shown in [\[link\]](#), it is given kinetic energy. The process is analogous to an object being accelerated by a gravitational field. It is as if the charge is going down an electrical hill where its electric potential energy is converted to kinetic energy. Let us explore the work done on a charge q by the electric field in this process, so that we may develop a definition of electric potential energy.



A charge accelerated by an electric field is analogous to a mass going down a hill. In both cases potential energy is converted to another form. Work is done by a force, but since this force

is conservative, we
can write
 $W = -\Delta\text{PE}$.

The electrostatic or Coulomb force is conservative, which means that the work done on q is independent of the path taken. This is exactly analogous to the gravitational force in the absence of dissipative forces such as friction. When a force is conservative, it is possible to define a potential energy associated with the force, and it is usually easier to deal with the potential energy (because it depends only on position) than to calculate the work directly.

We use the letters PE to denote electric potential energy, which has units of joules (J). The change in potential energy, ΔPE , is crucial, since the work done by a conservative force is the negative of the change in potential energy; that is, $W = -\Delta\text{PE}$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE . There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Note:

Potential Energy

$W = -\Delta\text{PE}$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE . There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation

without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of voltage (related to electric potential energy) than to deal with the Coulomb force directly.

Calculating the work directly is generally difficult, since $W = Fd \cos \theta$ and the direction and magnitude of F can be complex for multiple charges, for odd-shaped objects, and along arbitrary paths. But we do know that, since $F = qE$, the work, and hence ΔPE , is proportional to the test charge q . To have a physical quantity that is independent of test charge, we define **electric potential** V (or simply potential, since electric is understood) to be the potential energy per unit charge:

Equation:

$$V = \frac{PE}{q}.$$

Note:

Electric Potential

This is the electric potential energy per unit charge.

Equation:

$$V = \frac{PE}{q}$$

Since PE is proportional to q , the dependence on q cancels. Thus V does not depend on q . The change in potential energy ΔPE is crucial, and so we are concerned with the difference in potential or potential difference ΔV between two points, where

Equation:

$$\Delta V = V_B - V_A = \frac{\Delta PE}{q}.$$

The **potential difference** between points A and B, $V_B - V_A$, is thus defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

Equation:

$$1 \text{ V} = 1 \frac{\text{J}}{\text{C}}$$

Note:

Potential Difference

The potential difference between points A and B, $V_B - V_A$, is defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

Equation:

$$1 \text{ V} = 1 \frac{\text{J}}{\text{C}}$$

The familiar term **voltage** is the common name for potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor.

In summary, the relationship between potential difference (or voltage) and electrical potential energy is given by

Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

Note:**Potential Difference and Electrical Potential Energy**

The relationship between potential difference (or voltage) and electrical potential energy is given by

Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

The second equation is equivalent to the first.

Voltage is not the same as energy. Voltage is the energy per unit charge. Thus a motorcycle battery and a car battery can both have the same voltage (more precisely, the same potential difference between battery terminals), yet one stores much more energy than the other since $\Delta \text{PE} = q\Delta V$. The car battery can move more charge than the motorcycle battery, although both are 12 V batteries.

Example:**Calculating Energy**

Suppose you have a 12.0 V motorcycle battery that can move 5000 C of charge, and a 12.0 V car battery that can move 60,000 C of charge. How much energy does each deliver? (Assume that the numerical value of each charge is accurate to three significant figures.)

Strategy

To say we have a 12.0 V battery means that its terminals have a 12.0 V potential difference. When such a battery moves charge, it puts the charge

through a potential difference of 12.0 V, and the charge is given a change in potential energy equal to $\Delta PE = q\Delta V$.

So to find the energy output, we multiply the charge moved by the potential difference.

Solution

For the motorcycle battery, $q = 5000 \text{ C}$ and $\Delta V = 12.0 \text{ V}$. The total energy delivered by the motorcycle battery is

Equation:

$$\begin{aligned}\Delta PE_{\text{cycle}} &= (5000 \text{ C})(12.0 \text{ V}) \\ &= (5000 \text{ C})(12.0 \text{ J/C}) \\ &= 6.00 \times 10^4 \text{ J.}\end{aligned}$$

Similarly, for the car battery, $q = 60,000 \text{ C}$ and

Equation:

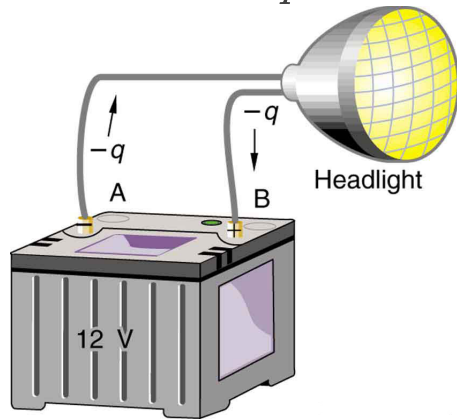
$$\begin{aligned}\Delta PE_{\text{car}} &= (60,000 \text{ C})(12.0 \text{ V}) \\ &= 7.20 \times 10^5 \text{ J.}\end{aligned}$$

Discussion

While voltage and energy are related, they are not the same thing. The voltages of the batteries are identical, but the energy supplied by each is quite different. Note also that as a battery is discharged, some of its energy is used internally and its terminal voltage drops, such as when headlights dim because of a low car battery. The energy supplied by the battery is still calculated as in this example, but not all of the energy is available for external use.

Note that the energies calculated in the previous example are absolute values. The change in potential energy for the battery is negative, since it loses energy. These batteries, like many electrical systems, actually move negative charge—electrons in particular. The batteries repel electrons from their negative terminals (A) through whatever circuitry is involved and attract them to their positive terminals (B) as shown in [\[link\]](#). The change in potential is $\Delta V = V_B - V_A = +12 \text{ V}$ and the charge q is negative, so that

$\Delta PE = q\Delta V$ is negative, meaning the potential energy of the battery has decreased when q has moved from A to B.



A battery moves negative charge from its negative terminal through a headlight to its positive terminal.

Appropriate combinations of chemicals in the battery separate charges so that the negative terminal has an excess of negative charge, which is repelled by it and attracted to the excess positive charge on the other terminal. In terms of potential, the positive terminal is at a higher voltage than the negative. Inside the battery, both positive and negative charges move.

Example:**How Many Electrons Move through a Headlight Each Second?**

When a 12.0 V car battery runs a single 30.0 W headlight, how many electrons pass through it each second?

Strategy

To find the number of electrons, we must first find the charge that moved in 1.00 s. The charge moved is related to voltage and energy through the equation $\Delta PE = q\Delta V$. A 30.0 W lamp uses 30.0 joules per second. Since the battery loses energy, we have $\Delta PE = -30.0 \text{ J}$ and, since the electrons are going from the negative terminal to the positive, we see that $\Delta V = +12.0 \text{ V}$.

Solution

To find the charge q moved, we solve the equation $\Delta PE = q\Delta V$:

Equation:

$$q = \frac{\Delta PE}{\Delta V}.$$

Entering the values for ΔPE and ΔV , we get

Equation:

$$q = \frac{-30.0 \text{ J}}{+12.0 \text{ V}} = \frac{-30.0 \text{ J}}{+12.0 \text{ J/C}} = -2.50 \text{ C}.$$

The number of electrons n_e is the total charge divided by the charge per electron. That is,

Equation:

$$n_e = \frac{-2.50 \text{ C}}{-1.60 \times 10^{-19} \text{ C/e}^-} = 1.56 \times 10^{19} \text{ electrons}.$$

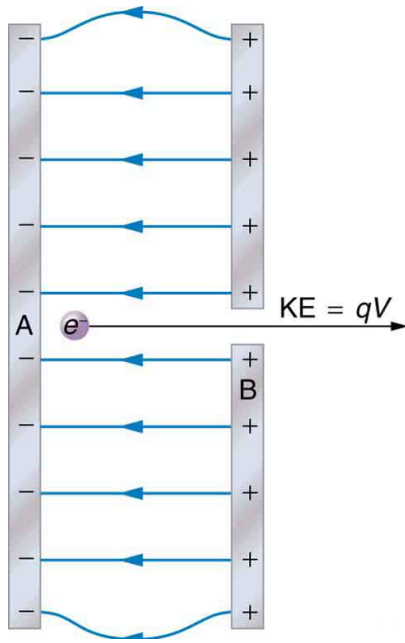
Discussion

This is a very large number. It is no wonder that we do not ordinarily observe individual electrons with so many being present in ordinary

systems. In fact, electricity had been in use for many decades before it was determined that the moving charges in many circumstances were negative. Positive charge moving in the opposite direction of negative charge often produces identical effects; this makes it difficult to determine which is moving or whether both are moving.

The Electron Volt

The energy per electron is very small in macroscopic situations like that in the previous example—a tiny fraction of a joule. But on a submicroscopic scale, such energy per particle (electron, proton, or ion) can be of great importance. For example, even a tiny fraction of a joule can be great enough for these particles to destroy organic molecules and harm living tissue. The particle may do its damage by direct collision, or it may create harmful x rays, which can also inflict damage. It is useful to have an energy unit related to submicroscopic effects. [\[link\]](#) shows a situation related to the definition of such an energy unit. An electron is accelerated between two charged metal plates as it might be in an old-model television tube or oscilloscope. The electron is given kinetic energy that is later converted to another form—light in the television tube, for example. (Note that downhill for the electron is uphill for a positive charge.) Since energy is related to voltage by $\Delta PE = q\Delta V$, we can think of the joule as a coulomb-volt.



A typical electron gun accelerates electrons using a potential difference between two metal plates. The energy of the electron in electron volts is numerically the same as the voltage between the plates.

For example, a 5000 V potential difference produces 5000 eV electrons.

On the submicroscopic scale, it is more convenient to define an energy unit called the **electron volt** (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form, **Equation:**

$$\begin{aligned}
 1 \text{ eV} &= (1.60 \times 10^{-19} \text{ C})(1 \text{ V}) = (1.60 \times 10^{-19} \text{ C})(1 \text{ J/C}) \\
 &= 1.60 \times 10^{-19} \text{ J}.
 \end{aligned}$$

Note:

Electron Volt

On the submicroscopic scale, it is more convenient to define an energy unit called the electron volt (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

Equation:

$$\begin{aligned}
 1 \text{ eV} &= (1.60 \times 10^{-19} \text{ C})(1 \text{ V}) = (1.60 \times 10^{-19} \text{ C})(1 \text{ J/C}) \\
 &= 1.60 \times 10^{-19} \text{ J}.
 \end{aligned}$$

An electron accelerated through a potential difference of 1 V is given an energy of 1 eV. It follows that an electron accelerated through 50 V is given 50 eV. A potential difference of 100,000 V (100 kV) will give an electron an energy of 100,000 eV (100 keV), and so on. Similarly, an ion with a double positive charge accelerated through 100 V will be given 200 eV of energy. These simple relationships between accelerating voltage and particle charges make the electron volt a simple and convenient energy unit in such circumstances.

Note:

Connections: Energy Units

The electron volt (eV) is the most common energy unit for submicroscopic processes. This will be particularly noticeable in the chapters on modern physics. Energy is so important to so many subjects that there is a tendency to define a special energy unit for each major topic. There are, for example,

calories for food energy, kilowatt-hours for electrical energy, and therms for natural gas energy.

The electron volt is commonly employed in submicroscopic processes—chemical valence energies and molecular and nuclear binding energies are among the quantities often expressed in electron volts. For example, about 5 eV of energy is required to break up certain organic molecules. If a proton is accelerated from rest through a potential difference of 30 kV, it is given an energy of 30 keV (30,000 eV) and it can break up as many as 6000 of these molecules ($30,000 \text{ eV} \div 5 \text{ eV per molecule} = 6000 \text{ molecules}$). Nuclear decay energies are on the order of 1 MeV (1,000,000 eV) per event and can, thus, produce significant biological damage.

Conservation of Energy

The total energy of a system is conserved if there is no net addition (or subtraction) of work or heat transfer. For conservative forces, such as the electrostatic force, conservation of energy states that mechanical energy is a constant.

Mechanical energy is the sum of the kinetic energy and potential energy of a system; that is, $KE + PE = \text{constant}$. A loss of PE of a charged particle becomes an increase in its KE. Here PE is the electric potential energy. Conservation of energy is stated in equation form as

Equation:

$$KE + PE = \text{constant}$$

or

Equation:

$$KE_i + PE_i = KE_f + PE_f,$$

where i and f stand for initial and final conditions. As we have found many times before, considering energy can give us insights and facilitate problem solving.

Example:

Electrical Potential Energy Converted to Kinetic Energy

Calculate the final speed of a free electron accelerated from rest through a potential difference of 100 V. (Assume that this numerical value is accurate to three significant figures.)

Strategy

We have a system with only conservative forces. Assuming the electron is accelerated in a vacuum, and neglecting the gravitational force (we will check on this assumption later), all of the electrical potential energy is converted into kinetic energy. We can identify the initial and final forms of energy to be $KE_i = 0$, $KE_f = \frac{1}{2}mv^2$, $PE_i = qV$, and $PE_f = 0$.

Solution

Conservation of energy states that

Equation:

$$KE_i + PE_i = KE_f + PE_f.$$

Entering the forms identified above, we obtain

Equation:

$$qV = \frac{mv^2}{2}.$$

We solve this for v :

Equation:

$$v = \sqrt{\frac{2qV}{m}}.$$

Entering values for q , V , and m gives

Equation:

$$\begin{aligned}
 v &= \sqrt{\frac{2(-1.60 \times 10^{-19} \text{ C})(-100 \text{ J/C})}{9.11 \times 10^{-31} \text{ kg}}} \\
 &= 5.93 \times 10^6 \text{ m/s.}
 \end{aligned}$$

Discussion

Note that both the charge and the initial voltage are negative, as in [\[link\]](#). From the discussions in [Electric Charge and Electric Field](#), we know that electrostatic forces on small particles are generally very large compared with the gravitational force. The large final speed confirms that the gravitational force is indeed negligible here. The large speed also indicates how easy it is to accelerate electrons with small voltages because of their very small mass. Voltages much higher than the 100 V in this problem are typically used in electron guns. Those higher voltages produce electron speeds so great that relativistic effects must be taken into account. That is why a low voltage is considered (accurately) in this example.

Section Summary

- Electric potential is potential energy per unit charge.
- The potential difference between points A and B, $V_B - V_A$, defined to be the change in potential energy of a charge q moved from A to B, is equal to the change in potential energy divided by the charge. Potential difference is commonly called voltage, represented by the symbol ΔV .

Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

- An electron volt is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

Equation:

$$\begin{aligned}
 1 \text{ eV} &= (1.60 \times 10^{-19} \text{ C})(1 \text{ V}) = (1.60 \times 10^{-19} \text{ C})(1 \text{ J/C}) \\
 &= 1.60 \times 10^{-19} \text{ J.}
 \end{aligned}$$

- Mechanical energy is the sum of the kinetic energy and potential energy of a system, that is, $KE + PE$. This sum is a constant.

Conceptual Questions

Exercise:

Problem:

Voltage is the common word for potential difference. Which term is more descriptive, voltage or potential difference?

Exercise:

Problem:

If the voltage between two points is zero, can a test charge be moved between them with zero net work being done? Can this necessarily be done without exerting a force? Explain.

Exercise:

Problem:

What is the relationship between voltage and energy? More precisely, what is the relationship between potential difference and electric potential energy?

Exercise:

Problem: Voltages are always measured between two points. Why?

Exercise:

Problem:

How are units of volts and electron volts related? How do they differ?

Problems & Exercises

Exercise:

Problem:

Find the ratio of speeds of an electron and a negative hydrogen ion (one having an extra electron) accelerated through the same voltage, assuming non-relativistic final speeds. Take the mass of the hydrogen ion to be 1.67×10^{-27} kg.

Solution:

42.8

Exercise:**Problem:**

An evacuated tube uses an accelerating voltage of 40 kV to accelerate electrons to hit a copper plate and produce x rays. Non-relativistically, what would be the maximum speed of these electrons?

Exercise:**Problem:**

A bare helium nucleus has two positive charges and a mass of 6.64×10^{-27} kg. (a) Calculate its kinetic energy in joules at 2.00% of the speed of light. (b) What is this in electron volts? (c) What voltage would be needed to obtain this energy?

Exercise:**Problem: Integrated Concepts**

Singly charged gas ions are accelerated from rest through a voltage of 13.0 V. At what temperature will the average kinetic energy of gas molecules be the same as that given these ions?

Solution:

1.00×10^5 K

Exercise:**Problem: Integrated Concepts**

The temperature near the center of the Sun is thought to be 15 million degrees Celsius (1.5×10^7 °C). Through what voltage must a singly charged ion be accelerated to have the same energy as the average kinetic energy of ions at this temperature?

Exercise:**Problem: Integrated Concepts**

(a) What is the average power output of a heart defibrillator that dissipates 400 J of energy in 10.0 ms? (b) Considering the high-power output, why doesn't the defibrillator produce serious burns?

Solution:

(a) 4×10^4 W

(b) A defibrillator does not cause serious burns because the skin conducts electricity well at high voltages, like those used in defibrillators. The gel used aids in the transfer of energy to the body, and the skin doesn't absorb the energy, but rather lets it pass through to the heart.

Exercise:**Problem: Integrated Concepts**

A lightning bolt strikes a tree, moving 20.0 C of charge through a potential difference of 1.00×10^2 MV. (a) What energy was dissipated? (b) What mass of water could be raised from 15°C to the boiling point and then boiled by this energy? (c) Discuss the damage that could be caused to the tree by the expansion of the boiling steam.

Exercise:

Problem: Integrated Concepts

A 12.0 V battery-operated bottle warmer heats 50.0 g of glass, 2.50×10^2 g of baby formula, and 2.00×10^2 g of aluminum from 20.0°C to 90.0°C. (a) How much charge is moved by the battery? (b) How many electrons per second flow if it takes 5.00 min to warm the formula? (Hint: Assume that the specific heat of baby formula is about the same as the specific heat of water.)

Solution:

(a) 7.40×10^3 C

(b) 1.54×10^{20} electrons per second

Exercise:**Problem: Integrated Concepts**

A battery-operated car utilizes a 12.0 V system. Find the charge the batteries must be able to move in order to accelerate the 750 kg car from rest to 25.0 m/s, make it climb a 2.00×10^2 m high hill, and then cause it to travel at a constant 25.0 m/s by exerting a 5.00×10^2 N force for an hour.

Solution:

3.89×10^6 C

Exercise:**Problem: Integrated Concepts**

Fusion probability is greatly enhanced when appropriate nuclei are brought close together, but mutual Coulomb repulsion must be overcome. This can be done using the kinetic energy of high-temperature gas ions or by accelerating the nuclei toward one another.

(a) Calculate the potential energy of two singly charged nuclei separated by 1.00×10^{-12} m by finding the voltage of one at that distance and multiplying by the charge of the other. (b) At what temperature will atoms of a gas have an average kinetic energy equal to this needed electrical potential energy?

Exercise:

Problem: Unreasonable Results

(a) Find the voltage near a 10.0 cm diameter metal sphere that has 8.00 C of excess positive charge on it. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) 1.44×10^{12} V

(b) This voltage is very high. A 10.0 cm diameter sphere could never maintain this voltage; it would discharge.

(c) An 8.00 C charge is more charge than can reasonably be accumulated on a sphere of that size.

Exercise:

Problem: Construct Your Own Problem

Consider a battery used to supply energy to a cellular phone. Construct a problem in which you determine the energy that must be supplied by the battery, and then calculate the amount of charge it must be able to move in order to supply this energy. Among the things to be considered are the energy needs and battery voltage. You may need to look ahead to interpret manufacturer's battery ratings in ampere-hours as energy in joules.

Glossary

electric potential

potential energy per unit charge

potential difference (or voltage)

change in potential energy of a charge moved from one point to another, divided by the charge; units of potential difference are joules per coulomb, known as volt

electron volt

the energy given to a fundamental charge accelerated through a potential difference of one volt

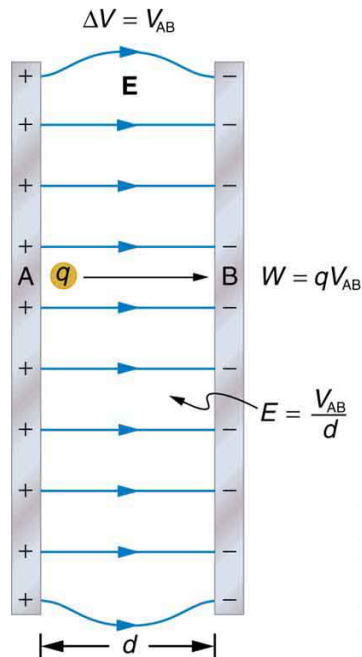
mechanical energy

sum of the kinetic energy and potential energy of a system; this sum is a constant

Electric Potential in a Uniform Electric Field

- Describe the relationship between voltage and electric field.
- Derive an expression for the electric potential and electric field.
- Calculate electric field strength given distance and voltage.

In the previous section, we explored the relationship between voltage and energy. In this section, we will explore the relationship between voltage and electric field. For example, a uniform electric field \mathbf{E} is produced by placing a potential difference (or voltage) ΔV across two parallel metal plates, labeled A and B. (See [\[link\]](#).) Examining this will tell us what voltage is needed to produce a certain electric field strength; it will also reveal a more fundamental relationship between electric potential and electric field. From a physicist's point of view, either ΔV or \mathbf{E} can be used to describe any charge distribution. ΔV is most closely tied to energy, whereas \mathbf{E} is most closely related to force. ΔV is a **scalar** quantity and has no direction, while \mathbf{E} is a **vector** quantity, having both magnitude and direction. (Note that the magnitude of the electric field strength, a scalar quantity, is represented by E below.) The relationship between ΔV and \mathbf{E} is revealed by calculating the work done by the force in moving a charge from point A to point B. But, as noted in [Electric Potential Energy: Potential Difference](#), this is complex for arbitrary charge distributions, requiring calculus. We therefore look at a uniform electric field as an interesting special case.



The relationship between V and E for parallel conducting plates is $E = V/d$. (Note that $\Delta V = V_{AB}$ in magnitude. For a charge that is moved from plate A at higher potential to plate B at lower potential, a minus sign needs to be included as follows:
 $-\Delta V = V_A - V_B = V_{AB}$
 . See the text for details.)

The work done by the electric field in [\[link\]](#) to move a positive charge q from A, the positive plate, higher potential, to B, the negative plate, lower potential, is

Equation:

$$W = -\Delta PE = -q\Delta V.$$

The potential difference between points A and B is

Equation:

$$-\Delta V = -(V_B - V_A) = V_A - V_B = V_{AB}.$$

Entering this into the expression for work yields

Equation:

$$W = qV_{AB}.$$

Work is $W = Fd \cos \theta$; here $\cos \theta = 1$, since the path is parallel to the field, and so $W = Fd$. Since $F = qE$, we see that $W = qEd$. Substituting this expression for work into the previous equation gives

Equation:

$$qEd = qV_{AB}.$$

The charge cancels, and so the voltage between points A and B is seen to be

Equation:

$$\left. \begin{array}{l} V_{AB} = Ed \\ E = \frac{V_{AB}}{d} \end{array} \right\} (\text{uniform } E - \text{field only}),$$

where d is the distance from A to B, or the distance between the plates in [\[link\]](#). Note that the above equation implies the units for electric field are volts per meter. We already know the units for electric field are newtons per coulomb; thus the following relation among units is valid:

Equation:

$$1 \text{ N/C} = 1 \text{ V/m}.$$

Note:

Voltage between Points A and B

Equation:

$$\left. \begin{array}{l} V_{AB} = Ed \\ E = \frac{V_{AB}}{d} \end{array} \right\} (\text{uniform } E - \text{field only}),$$

where d is the distance from A to B, or the distance between the plates.

Example:**What Is the Highest Voltage Possible between Two Plates?**

Dry air will support a maximum electric field strength of about $3.0 \times 10^6 \text{ V/m}$. Above that value, the field creates enough ionization in the air to make the air a conductor. This allows a discharge or spark that reduces the field. What, then, is the maximum voltage between two parallel conducting plates separated by 2.5 cm of dry air?

Strategy

We are given the maximum electric field E between the plates and the distance d between them. The equation $V_{AB} = Ed$ can thus be used to calculate the maximum voltage.

Solution

The potential difference or voltage between the plates is

Equation:

$$V_{AB} = Ed.$$

Entering the given values for E and d gives

Equation:

$$V_{AB} = (3.0 \times 10^6 \text{ V/m})(0.025 \text{ m}) = 7.5 \times 10^4 \text{ V}$$

or

Equation:

$$V_{AB} = 75 \text{ kV}.$$

(The answer is quoted to only two digits, since the maximum field strength is approximate.)

Discussion

One of the implications of this result is that it takes about 75 kV to make a spark jump across a 2.5 cm (1 in.) gap, or 150 kV for a 5 cm spark. This limits the voltages that can exist between conductors, perhaps on a power transmission line. A smaller voltage will cause a spark if there are points on the surface, since points create greater fields than smooth surfaces. Humid air breaks down at a lower field strength, meaning that a smaller voltage will make a spark jump through humid air. The largest voltages can be built up, say with static electricity, on dry days.



A spark chamber is used to trace the paths of high-energy particles. Ionization created by the particles as they pass through the gas between the plates allows a spark to jump. The sparks are

perpendicular to the plates, following electric field lines between them. The potential difference between adjacent plates is not high enough to cause sparks without the ionization produced by particles from accelerator experiments (or cosmic rays).
(credit: Daderot, Wikimedia Commons)

Example:**Field and Force inside an Electron Gun**

(a) An electron gun has parallel plates separated by 4.00 cm and gives electrons 25.0 keV of energy. What is the electric field strength between the plates? (b) What force would this field exert on a piece of plastic with a $0.500\text{ }\mu\text{C}$ charge that gets between the plates?

Strategy

Since the voltage and plate separation are given, the electric field strength can be calculated directly from the expression $E = \frac{V_{AB}}{d}$. Once the electric field strength is known, the force on a charge is found using $\mathbf{F} = q \mathbf{E}$. Since the electric field is in only one direction, we can write this equation in terms of the magnitudes, $F = q E$.

Solution for (a)

The expression for the magnitude of the electric field between two uniform metal plates is

Equation:

$$E = \frac{V_{AB}}{d}.$$

Since the electron is a single charge and is given 25.0 keV of energy, the potential difference must be 25.0 kV. Entering this value for V_{AB} and the plate separation of 0.0400 m, we obtain

Equation:

$$E = \frac{25.0 \text{ kV}}{0.0400 \text{ m}} = 6.25 \times 10^5 \text{ V/m}.$$

Solution for (b)

The magnitude of the force on a charge in an electric field is obtained from the equation

Equation:

$$F = qE.$$

Substituting known values gives

Equation:

$$F = (0.500 \times 10^{-6} \text{ C})(6.25 \times 10^5 \text{ V/m}) = 0.313 \text{ N}.$$

Discussion

Note that the units are newtons, since $1 \text{ V/m} = 1 \text{ N/C}$. The force on the charge is the same no matter where the charge is located between the plates. This is because the electric field is uniform between the plates.

In more general situations, regardless of whether the electric field is uniform, it points in the direction of decreasing potential, because the force on a positive charge is in the direction of \mathbf{E} and also in the direction of lower potential V . Furthermore, the magnitude of \mathbf{E} equals the rate of decrease of V with distance. The faster V decreases over distance, the

greater the electric field. In equation form, the general relationship between voltage and electric field is

Equation:

$$E = - \frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV , takes place. The minus sign tells us that \mathbf{E} points in the direction of decreasing potential. The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

Note:

Relationship between Voltage and Electric Field

In equation form, the general relationship between voltage and electric field is

Equation:

$$E = - \frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV , takes place. The minus sign tells us that \mathbf{E} points in the direction of decreasing potential. The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

For continually changing potentials, ΔV and Δs become infinitesimals and differential calculus must be employed to determine the electric field.

Section Summary

- The voltage between points A and B is

Equation:

$$\left. \begin{aligned} V_{AB} &= Ed \\ E &= \frac{V_{AB}}{d} \end{aligned} \right\} (\text{uniform } E - \text{field only}),$$

where d is the distance from A to B, or the distance between the plates.

- In equation form, the general relationship between voltage and electric field is

Equation:

$$E = - \frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV , takes place. The minus sign tells us that \mathbf{E} points in the direction of decreasing potential.) The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

Conceptual Questions

Exercise:

Problem:

Discuss how potential difference and electric field strength are related. Give an example.

Exercise:

Problem:

What is the strength of the electric field in a region where the electric potential is constant?

Exercise:

Problem:

Will a negative charge, initially at rest, move toward higher or lower potential? Explain why.

Problems & Exercises

Exercise:

Problem:

Show that units of V/m and N/C for electric field strength are indeed equivalent.

Exercise:

Problem:

What is the strength of the electric field between two parallel conducting plates separated by 1.00 cm and having a potential difference (voltage) between them of 1.50×10^4 V?

Exercise:

Problem:

The electric field strength between two parallel conducting plates separated by 4.00 cm is 7.50×10^4 V/m. (a) What is the potential difference between the plates? (b) The plate with the lowest potential is taken to be at zero volts. What is the potential 1.00 cm from that plate (and 3.00 cm from the other)?

Solution:

(a) 3.00 kV

(b) 750 V

Exercise:

Problem:

How far apart are two conducting plates that have an electric field strength of 4.50×10^3 V/m between them, if their potential difference is 15.0 kV?

Exercise:

Problem:

(a) Will the electric field strength between two parallel conducting plates exceed the breakdown strength for air ($3.0 \times 10^6 \text{ V/m}$) if the plates are separated by 2.00 mm and a potential difference of $5.0 \times 10^3 \text{ V}$ is applied? (b) How close together can the plates be with this applied voltage?

Solution:

(a) No. The electric field strength between the plates is $2.5 \times 10^6 \text{ V/m}$, which is lower than the breakdown strength for air ($3.0 \times 10^6 \text{ V/m}$).

(b) 1.7 mm

Exercise:**Problem:**

The voltage across a membrane forming a cell wall is 80.0 mV and the membrane is 9.00 nm thick. What is the electric field strength? (The value is surprisingly large, but correct. Membranes are discussed in [Capacitors and Dielectrics](#) and [Nerve Conduction—Electrocardiograms](#).) You may assume a uniform electric field.

Exercise:**Problem:**

Membrane walls of living cells have surprisingly large electric fields across them due to separation of ions. (Membranes are discussed in some detail in [Nerve Conduction—Electrocardiograms](#).) What is the voltage across an 8.00 nm-thick membrane if the electric field strength across it is 5.50 MV/m? You may assume a uniform electric field.

Solution:

44.0 mV

Exercise:**Problem:**

Two parallel conducting plates are separated by 10.0 cm, and one of them is taken to be at zero volts. (a) What is the electric field strength between them, if the potential 8.00 cm from the zero volt plate (and 2.00 cm from the other) is 450 V? (b) What is the voltage between the plates?

Exercise:**Problem:**

Find the maximum potential difference between two parallel conducting plates separated by 0.500 cm of air, given the maximum sustainable electric field strength in air to be $3.0 \times 10^6 \text{ V/m}$.

Solution:

15 kV

Exercise:**Problem:**

A doubly charged ion is accelerated to an energy of 32.0 keV by the electric field between two parallel conducting plates separated by 2.00 cm. What is the electric field strength between the plates?

Exercise:**Problem:**

An electron is to be accelerated in a uniform electric field having a strength of $2.00 \times 10^6 \text{ V/m}$. (a) What energy in keV is given to the electron if it is accelerated through 0.400 m? (b) Over what distance would it have to be accelerated to increase its energy by 50.0 GeV?

Solution:

(a) 800 KeV

(b) 25.0 km

Glossary

scalar

physical quantity with magnitude but no direction

vector

physical quantity with both magnitude and direction

Electrical Potential Due to a Point Charge

- Explain point charges and express the equation for electric potential of a point charge.
- Distinguish between electric potential and electric field.
- Determine the electric potential of a point charge given charge and distance.

Point charges, such as electrons, are among the fundamental building blocks of matter. Furthermore, spherical charge distributions (like on a metal sphere) create external electric fields exactly like a point charge. The electric potential due to a point charge is, thus, a case we need to consider. Using calculus to find the work needed to move a test charge q from a large distance away to a distance of r from a point charge Q , and noting the connection between work and potential ($W = -q\Delta V$), it can be shown that the *electric potential V of a point charge* is

Equation:

$$V = \frac{kQ}{r} \text{ (Point Charge),}$$

where k is a constant equal to $9.0 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$.

Note:

Electric Potential V of a Point Charge

The electric potential V of a point charge is given by

Equation:

$$V = \frac{kQ}{r} \text{ (Point Charge).}$$

The potential at infinity is chosen to be zero. Thus V for a point charge decreases with distance, whereas \mathbf{E} for a point charge decreases with distance squared:

Equation:

$$E = \frac{F}{q} = \frac{kQ}{r^2}.$$

Recall that the electric potential V is a scalar and has no direction, whereas the electric field \mathbf{E} is a vector. To find the voltage due to a combination of point charges, you add the individual voltages as numbers. To find the total electric field, you must add the individual fields as *vectors*, taking magnitude and direction into account. This is consistent with the fact that V is closely associated with energy, a scalar, whereas \mathbf{E} is closely associated with force, a vector.

Example:

What Voltage Is Produced by a Small Charge on a Metal Sphere?

Charges in static electricity are typically in the nanocoulomb (nC) to microcoulomb (μC) range. What is the voltage 5.00 cm away from the center of a 1-cm diameter metal sphere that has a -3.00 nC static charge?

Strategy

As we have discussed in [Electric Charge and Electric Field](#), charge on a metal sphere spreads out uniformly and produces a field like that of a point charge located at its center. Thus we can find the voltage using the equation $V = kQ/r$.

Solution

Entering known values into the expression for the potential of a point charge, we obtain

Equation:

$$\begin{aligned}
 V &= k \frac{Q}{r} \\
 &= (8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \left(\frac{-3.00 \times 10^{-9} \text{ C}}{5.00 \times 10^{-2} \text{ m}} \right) \\
 &= -539 \text{ V}.
 \end{aligned}$$

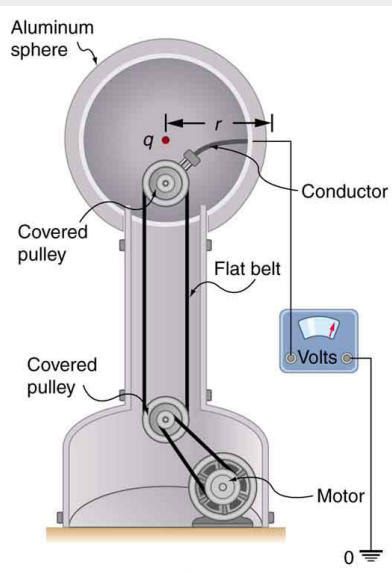
Discussion

The negative value for voltage means a positive charge would be attracted from a larger distance, since the potential is lower (more negative) than at larger distances. Conversely, a negative charge would be repelled, as expected.

Example:

What Is the Excess Charge on a Van de Graaff Generator

A demonstration Van de Graaff generator has a 25.0 cm diameter metal sphere that produces a voltage of 100 kV near its surface. (See [link](#).) What excess charge resides on the sphere? (Assume that each numerical value here is shown with three significant figures.)



The voltage of this demonstration Van de Graaff generator is measured

between the charged sphere and ground. Earth's potential is taken to be zero as a reference. The potential of the charged conducting sphere is the same as that of an equal point charge at its center.

Strategy

The potential on the surface will be the same as that of a point charge at the center of the sphere, 12.5 cm away. (The radius of the sphere is 12.5 cm.)

We can thus determine the excess charge using the equation

Equation:

$$V = \frac{kQ}{r}.$$

Solution

Solving for Q and entering known values gives

Equation:

$$\begin{aligned} Q &= \frac{rV}{k} \\ &= \frac{(0.125 \text{ m})(100 \times 10^3 \text{ V})}{8.99 \times 10^9 \text{ N} \cdot \text{m}^2 / \text{C}^2} \\ &= 1.39 \times 10^{-6} \text{ C} = 1.39 \text{ } \mu\text{C}. \end{aligned}$$

Discussion

This is a relatively small charge, but it produces a rather large voltage. We have another indication here that it is difficult to store isolated charges.

The voltages in both of these examples could be measured with a meter that compares the measured potential with ground potential. Ground potential is often taken to be zero (instead of taking the potential at infinity to be zero). It is the potential difference between two points that is of importance, and very often there is a tacit assumption that some reference point, such as Earth or a very distant point, is at zero potential. As noted in [Electric Potential Energy: Potential Difference](#), this is analogous to taking sea level as $h = 0$ when considering gravitational potential energy, $PE_g = mgh$.

Section Summary

- Electric potential of a point charge is $V = kQ/r$.
- Electric potential is a scalar, and electric field is a vector. Addition of voltages as numbers gives the voltage due to a combination of point charges, whereas addition of individual fields as vectors gives the total electric field.

Conceptual Questions

Exercise:

Problem:

In what region of space is the potential due to a uniformly charged sphere the same as that of a point charge? In what region does it differ from that of a point charge?

Exercise:

Problem:

Can the potential of a non-uniformly charged sphere be the same as that of a point charge? Explain.

Problems & Exercises

Exercise:

Problem:

A 0.500 cm diameter plastic sphere, used in a static electricity demonstration, has a uniformly distributed 40.0 pC charge on its surface. What is the potential near its surface?

Solution:

144 V

Exercise:**Problem:**

What is the potential 0.530×10^{-10} m from a proton (the average distance between the proton and electron in a hydrogen atom)?

Exercise:**Problem:**

(a) A sphere has a surface uniformly charged with 1.00 C. At what distance from its center is the potential 5.00 MV? (b) What does your answer imply about the practical aspect of isolating such a large charge?

Solution:

(a) 1.80 km

(b) A charge of 1 C is a very large amount of charge; a sphere of radius 1.80 km is not practical.

Exercise:**Problem:**

How far from a 1.00 μC point charge will the potential be 100 V? At what distance will it be 2.00×10^2 V?

Exercise:

Problem:

What are the sign and magnitude of a point charge that produces a potential of -2.00 V at a distance of 1.00 mm ?

Solution:

$$-2.22 \times 10^{-13}\text{ C}$$

Exercise:**Problem:**

If the potential due to a point charge is $5.00 \times 10^2\text{ V}$ at a distance of 15.0 m , what are the sign and magnitude of the charge?

Exercise:**Problem:**

In nuclear fission, a nucleus splits roughly in half. (a) What is the potential $2.00 \times 10^{-14}\text{ m}$ from a fragment that has 46 protons in it? (b) What is the potential energy in MeV of a similarly charged fragment at this distance?

Solution:

(a) $3.31 \times 10^6\text{ V}$

(b) 152 MeV

Exercise:**Problem:**

A research Van de Graaff generator has a 2.00-m -diameter metal sphere with a charge of 5.00 mC on it. (a) What is the potential near its surface? (b) At what distance from its center is the potential 1.00 MV ? (c) An oxygen atom with three missing electrons is released near the Van de Graaff generator. What is its energy in MeV at this distance?

Exercise:**Problem:**

An electrostatic paint sprayer has a 0.200-m-diameter metal sphere at a potential of 25.0 kV that repels paint droplets onto a grounded object.

(a) What charge is on the sphere? (b) What charge must a 0.100-mg drop of paint have to arrive at the object with a speed of 10.0 m/s?

Solution:

(a) $2.78 \times 10^{-7} \text{ C}$

(b) $2.00 \times 10^{-10} \text{ C}$

Exercise:**Problem:**

In one of the classic nuclear physics experiments at the beginning of the 20th century, an alpha particle was accelerated toward a gold nucleus, and its path was substantially deflected by the Coulomb interaction. If the energy of the doubly charged alpha nucleus was 5.00 MeV, how close to the gold nucleus (79 protons) could it come before being deflected?

Exercise:**Problem:**

(a) What is the potential between two points situated 10 cm and 20 cm from a $3.0 \mu\text{C}$ point charge? (b) To what location should the point at 20 cm be moved to increase this potential difference by a factor of two?

Exercise:**Problem: Unreasonable Results**

(a) What is the final speed of an electron accelerated from rest through a voltage of 25.0 MV by a negatively charged Van de Graaff terminal?

(b) What is unreasonable about this result?

(c) Which assumptions are responsible?

Solution:

(a) $2.96 \times 10^9 \text{ m/s}$

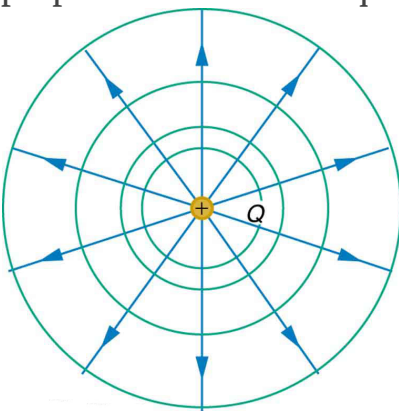
(b) This velocity is far too great. It is faster than the speed of light.

(c) The assumption that the speed of the electron is far less than that of light and that the problem does not require a relativistic treatment produces an answer greater than the speed of light.

Equipotential Lines

- Explain equipotential lines and equipotential surfaces.
- Describe the action of grounding an electrical appliance.
- Compare electric field and equipotential lines.

We can represent electric potentials (voltages) pictorially, just as we drew pictures to illustrate electric fields. Of course, the two are related. Consider [\[link\]](#), which shows an isolated positive point charge and its electric field lines. Electric field lines radiate out from a positive charge and terminate on negative charges. While we use blue arrows to represent the magnitude and direction of the electric field, we use green lines to represent places where the electric potential is constant. These are called **equipotential lines** in two dimensions, or *equipotential surfaces* in three dimensions. The term *equipotential* is also used as a noun, referring to an equipotential line or surface. The potential for a point charge is the same anywhere on an imaginary sphere of radius r surrounding the charge. This is true since the potential for a point charge is given by $V = kQ/r$ and, thus, has the same value at any point that is a given distance r from the charge. An equipotential sphere is a circle in the two-dimensional view of [\[link\]](#). Since the electric field lines point radially away from the charge, they are perpendicular to the equipotential lines.



An isolated point charge Q with its electric field lines in blue and equipotential lines

in green. The potential is the same along each equipotential line, meaning that no work is required to move a charge anywhere along one of those lines. Work is needed to move a charge from one equipotential line to another. Equipotential lines are perpendicular to electric field lines in every case.

It is important to note that *equipotential lines are always perpendicular to electric field lines*. No work is required to move a charge along an equipotential, since $\Delta V = 0$. Thus the work is

Equation:

$$W = -\Delta \text{PE} = -q\Delta V = 0.$$

Work is zero if force is perpendicular to motion. Force is in the same direction as \mathbf{E} , so that motion along an equipotential must be perpendicular to \mathbf{E} . More precisely, work is related to the electric field by

Equation:

$$W = Fd \cos \theta = qEd \cos \theta = 0.$$

Note that in the above equation, E and F symbolize the magnitudes of the electric field strength and force, respectively. Neither q nor \mathbf{E} nor d is zero, and so $\cos \theta$ must be 0, meaning θ must be 90° . In other words, motion along an equipotential is perpendicular to \mathbf{E} .

One of the rules for static electric fields and conductors is that the electric field must be perpendicular to the surface of any conductor. This implies that a *conductor is an equipotential surface in static situations*. There can be no voltage difference across the surface of a conductor, or charges will flow. One of the uses of this fact is that a conductor can be fixed at zero volts by connecting it to the earth with a good conductor—a process called **grounding**. Grounding can be a useful safety tool. For example, grounding the metal case of an electrical appliance ensures that it is at zero volts relative to the earth.

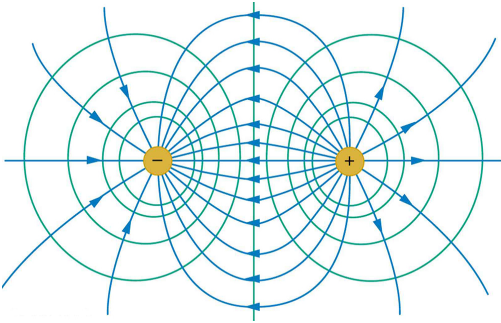
Note:

Grounding

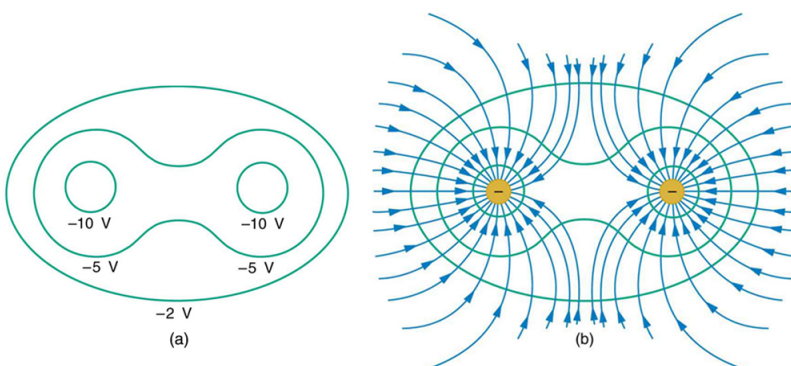
A conductor can be fixed at zero volts by connecting it to the earth with a good conductor—a process called grounding.

Because a conductor is an equipotential, it can replace any equipotential surface. For example, in [\[link\]](#) a charged spherical conductor can replace the point charge, and the electric field and potential surfaces outside of it will be unchanged, confirming the contention that a spherical charge distribution is equivalent to a point charge at its center.

[\[link\]](#) shows the electric field and equipotential lines for two equal and opposite charges. Given the electric field lines, the equipotential lines can be drawn simply by making them perpendicular to the electric field lines. Conversely, given the equipotential lines, as in [\[link\]\(a\)](#), the electric field lines can be drawn by making them perpendicular to the equipotentials, as in [\[link\]\(b\)](#).



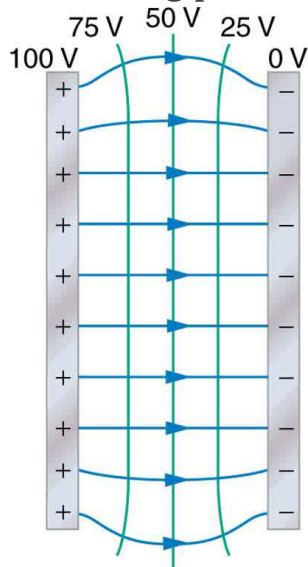
The electric field lines and equipotential lines for two equal but opposite charges. The equipotential lines can be drawn by making them perpendicular to the electric field lines, if those are known. Note that the potential is greatest (most positive) near the positive charge and least (most negative) near the negative charge.



(a) These equipotential lines might be measured with a voltmeter in a laboratory experiment. (b) The corresponding electric field lines are found by drawing them

perpendicular to the equipotentials. Note that these fields are consistent with two equal negative charges.

One of the most important cases is that of the familiar parallel conducting plates shown in [\[link\]](#). Between the plates, the equipotentials are evenly spaced and parallel. The same field could be maintained by placing conducting plates at the equipotential lines at the potentials shown.



The electric field and equipotential lines between two metal plates.

An important application of electric fields and equipotential lines involves the heart. The heart relies on electrical signals to maintain its rhythm. The movement of electrical signals causes the chambers of the heart to contract and relax. When a person has a heart attack, the movement of these electrical signals may be disturbed. An artificial pacemaker and a

defibrillator can be used to initiate the rhythm of electrical signals. The equipotential lines around the heart, the thoracic region, and the axis of the heart are useful ways of monitoring the structure and functions of the heart. An electrocardiogram (ECG) measures the small electric signals being generated during the activity of the heart. More about the relationship between electric fields and the heart is discussed in [Energy Stored in Capacitors](#).

Note:

PhET Explorations: Charges and Fields

Move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more. It's colorful, it's dynamic, it's free.

https://phet.colorado.edu/sims/html/charges-and-fields/latest/charges-and-fields_en.html

Section Summary

- An equipotential line is a line along which the electric potential is constant.
- An equipotential surface is a three-dimensional version of equipotential lines.
- Equipotential lines are always perpendicular to electric field lines.
- The process by which a conductor can be fixed at zero volts by connecting it to the earth with a good conductor is called grounding.

Conceptual Questions

Exercise:

Problem:

What is an equipotential line? What is an equipotential surface?

Exercise:

Problem:

Explain in your own words why equipotential lines and surfaces must be perpendicular to electric field lines.

Exercise:

Problem: Can different equipotential lines cross? Explain.

Problems & Exercises

Exercise:

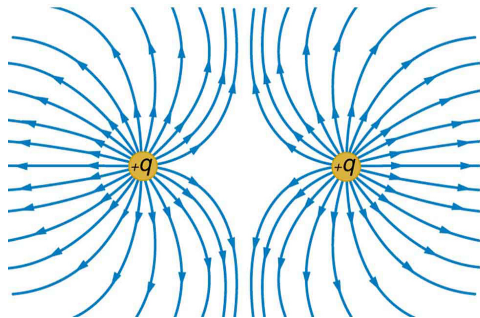
Problem:

(a) Sketch the equipotential lines near a point charge $+q$. Indicate the direction of increasing potential. (b) Do the same for a point charge $-3q$.

Exercise:

Problem:

Sketch the equipotential lines for the two equal positive charges shown in [\[link\]](#). Indicate the direction of increasing potential.



The electric field near two equal positive charges is directed away from each of the charges.

Exercise:

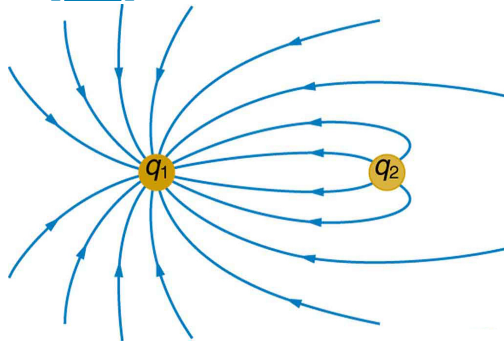
Problem:

[\[link\]](#) shows the electric field lines near two charges q_1 and q_2 , the first having a magnitude four times that of the second. Sketch the equipotential lines for these two charges, and indicate the direction of increasing potential.

Exercise:

Problem:

Sketch the equipotential lines a long distance from the charges shown in [\[link\]](#). Indicate the direction of increasing potential.



The electric field near
two charges.

Exercise:

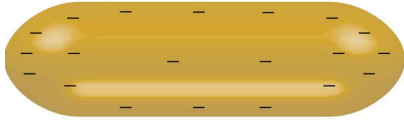
Problem:

Sketch the equipotential lines in the vicinity of two opposite charges, where the negative charge is three times as great in magnitude as the positive. See [\[link\]](#) for a similar situation. Indicate the direction of increasing potential.

Exercise:

Problem:

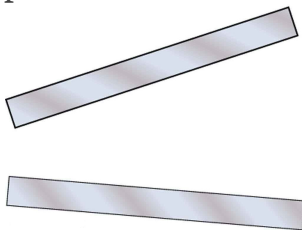
Sketch the equipotential lines in the vicinity of the negatively charged conductor in [\[link\]](#). How will these equipotentials look a long distance from the object?



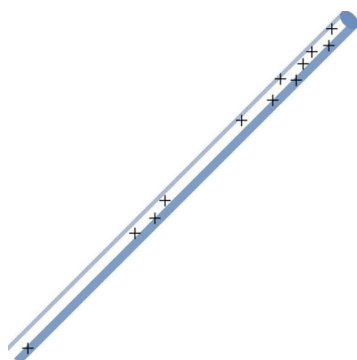
A negatively charged conductor.

Exercise:**Problem:**

Sketch the equipotential lines surrounding the two conducting plates shown in [\[link\]](#), given the top plate is positive and the bottom plate has an equal amount of negative charge. Be certain to indicate the distribution of charge on the plates. Is the field strongest where the plates are closest? Why should it be?

**Exercise:****Problem:**

(a) Sketch the electric field lines in the vicinity of the charged insulator in [\[link\]](#). Note its non-uniform charge distribution. (b) Sketch equipotential lines surrounding the insulator. Indicate the direction of increasing potential.



A charged insulating rod such as might be used in a classroom demonstration.

Exercise:

Problem:

The naturally occurring charge on the ground on a fine day out in the open country is -1.00 nC/m^2 . (a) What is the electric field relative to ground at a height of 3.00 m? (b) Calculate the electric potential at this height. (c) Sketch electric field and equipotential lines for this scenario.

Exercise:

Problem:

The lesser electric ray (*Narcine bancroftii*) maintains an incredible charge on its head and a charge equal in magnitude but opposite in sign on its tail ([\[link\]](#)). (a) Sketch the equipotential lines surrounding the ray. (b) Sketch the equipotentials when the ray is near a ship with a conducting surface. (c) How could this charge distribution be of use to the ray?



Lesser electric ray (*Narcine bancroftii*) (credit: National Oceanic and Atmospheric Administration, NOAA's Fisheries Collection).

Glossary

equipotential line

a line along which the electric potential is constant

grounding

fixing a conductor at zero volts by connecting it to the earth or ground

Introduction to Electric Current, Resistance, and Ohm's Law

class="introduction"

Electric energy in massive quantities is transmitted from this hydroelectric facility, the Srisailem power station located along the Krishna River in India, by the movement of charge—that is, by electric current.
(credit: Chintohere, Wikimedia Commons)



The flicker of numbers on a handheld calculator, nerve impulses carrying signals of vision to the brain, an ultrasound device sending a signal to a computer screen, the brain sending a message for a baby to twitch its toes, an electric train pulling its load over a mountain pass, a hydroelectric plant sending energy to metropolitan and rural users—these and many other examples of electricity involve *electric current, the movement of charge*. Humankind has indeed harnessed electricity, the basis of technology, to improve our quality of life. Whereas the previous two chapters concentrated on static electricity and the fundamental force underlying its behavior, the next few chapters will be devoted to electric and magnetic phenomena involving current. In addition to exploring applications of electricity, we shall gain new insights into nature—in particular, the fact that all magnetism results from electric current.

Current

- Define electric current, ampere, and drift velocity
- Describe the direction of charge flow in conventional current.
- Use drift velocity to calculate current and vice versa.

Electric Current

Electric current is defined to be the rate at which charge flows. A large current, such as that used to start a truck engine, moves a large amount of charge in a small time, whereas a small current, such as that used to operate a hand-held calculator, moves a small amount of charge over a long period of time. In equation form, **electric current** I is defined to be

Equation:

$$I = \frac{\Delta Q}{\Delta t},$$

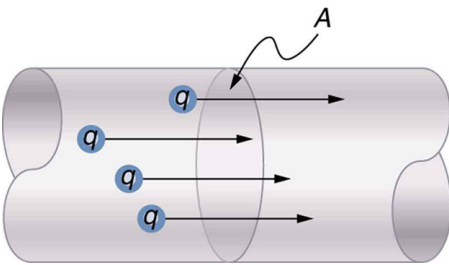
where ΔQ is the amount of charge passing through a given area in time Δt . (As in previous chapters, initial time is often taken to be zero, in which case $\Delta t = t$.) (See [\[link\]](#).) The SI unit for current is the **ampere** (A), named for the French physicist André-Marie Ampère (1775–1836). Since $I = \Delta Q / \Delta t$, we see that an ampere is one coulomb per second:

Equation:

$$1 \text{ A} = 1 \text{ C/s}$$

Not only are fuses and circuit breakers rated in amperes (or amps), so are many electrical appliances.

Current = flow of charge



The rate of flow of charge is current. An ampere is the flow of one coulomb through an area in one second.

Example:

Calculating Currents: Current in a Truck Battery and a Handheld Calculator

(a) What is the current involved when a truck battery sets in motion 720 C of charge in 4.00 s while starting an engine? (b) How long does it take 1.00 C of charge to flow through a handheld calculator if a 0.300-mA current is flowing?

Strategy

We can use the definition of current in the equation $I = \Delta Q / \Delta t$ to find the current in part (a), since charge and time are given. In part (b), we rearrange the definition of current and use the given values of charge and current to find the time required.

Solution for (a)

Entering the given values for charge and time into the definition of current gives

Equation:

$$\begin{aligned} I &= \frac{\Delta Q}{\Delta t} = \frac{720 \text{ C}}{4.00 \text{ s}} = 180 \text{ C/s} \\ &= 180 \text{ A.} \end{aligned}$$

Discussion for (a)

This large value for current illustrates the fact that a large charge is moved in a small amount of time. The currents in these “starter motors” are fairly large because large frictional forces need to be overcome when setting something in motion.

Solution for (b)

Solving the relationship $I = \Delta Q / \Delta t$ for time Δt , and entering the known values for charge and current gives

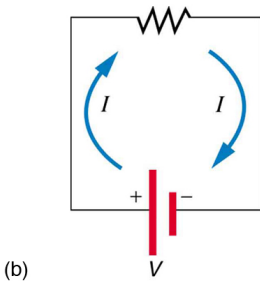
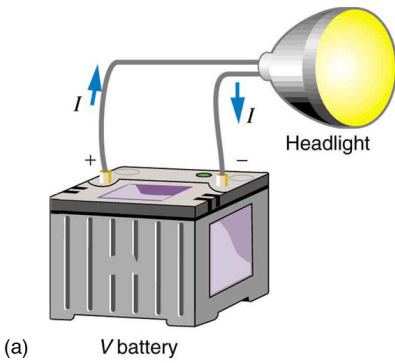
Equation:

$$\begin{aligned}\Delta t &= \frac{\Delta Q}{I} = \frac{1.00 \text{ C}}{0.300 \times 10^{-3} \text{ C/s}} \\ &= 3.33 \times 10^3 \text{ s.}\end{aligned}$$

Discussion for (b)

This time is slightly less than an hour. The small current used by the handheld calculator takes a much longer time to move a smaller charge than the large current of the truck starter. So why can we operate our calculators only seconds after turning them on? It's because calculators require very little energy. Such small current and energy demands allow handheld calculators to operate from solar cells or to get many hours of use out of small batteries. Remember, calculators do not have moving parts in the same way that a truck engine has with cylinders and pistons, so the technology requires smaller currents.

[\[link\]](#) shows a simple circuit and the standard schematic representation of a battery, conducting path, and load (a resistor). Schematics are very useful in visualizing the main features of a circuit. A single schematic can represent a wide variety of situations. The schematic in [\[link\]](#) (b), for example, can represent anything from a truck battery connected to a headlight lighting the street in front of the truck to a small battery connected to a penlight lighting a keyhole in a door. Such schematics are useful because the analysis is the same for a wide variety of situations. We need to understand a few schematics to apply the concepts and analysis to many more situations.



(a) A simple electric circuit. A closed path for current to flow through is supplied by conducting wires connecting a load to the terminals of a battery. (b) In this schematic, the battery is represented by the two parallel red lines, conducting wires are shown as straight lines, and the zigzag represents the load. The schematic represents a wide

variety of similar
circuits.

Note that the direction of current flow in [\[link\]](#) is from positive to negative. *The direction of conventional current is the direction that positive charge would flow.* Depending on the situation, positive charges, negative charges, or both may move. In metal wires, for example, current is carried by electrons—that is, negative charges move. In ionic solutions, such as salt water, both positive and negative charges move. This is also true in nerve cells. A Van de Graaff generator used for nuclear research can produce a current of pure positive charges, such as protons. [\[link\]](#) illustrates the movement of charged particles that compose a current. The fact that conventional current is taken to be in the direction that positive charge would flow can be traced back to American politician and scientist Benjamin Franklin in the 1700s. He named the type of charge associated with electrons negative, long before they were known to carry current in so many situations. Franklin, in fact, was totally unaware of the small-scale structure of electricity.

It is important to realize that there is an electric field in conductors responsible for producing the current, as illustrated in [\[link\]](#). Unlike static electricity, where a conductor in equilibrium cannot have an electric field in it, conductors carrying a current have an electric field and are not in static equilibrium. An electric field is needed to supply energy to move the charges.

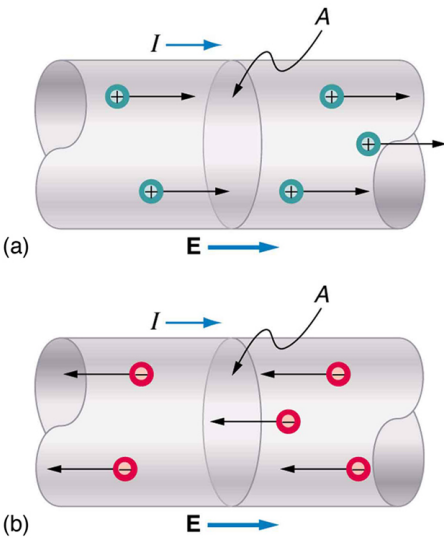
Note:

Making Connections: Take-Home Investigation—Electric Current Illustration

Find a straw and little peas that can move freely in the straw. Place the straw flat on a table and fill the straw with peas. When you pop one pea in at one end, a different pea should pop out the other end. This demonstration is an analogy for an electric current. Identify what compares

to the electrons and what compares to the supply of energy. What other analogies can you find for an electric current?

Note that the flow of peas is based on the peas physically bumping into each other; electrons flow due to mutually repulsive electrostatic forces.



Current I is the rate at which charge moves through an area A , such as the cross-section of a wire.

Conventional current is defined to move in the direction of the electric field. (a)

Positive charges move in the direction of the electric field and the same direction as conventional current.

(b) Negative charges move in the direction opposite to the electric field. Conventional

current is in the direction opposite to the movement of negative charge. The flow of electrons is sometimes referred to as electronic flow.

Example:**Calculating the Number of Electrons that Move through a Calculator**

If the 0.300-mA current through the calculator mentioned in the [\[link\]](#) example is carried by electrons, how many electrons per second pass through it?

Strategy

The current calculated in the previous example was defined for the flow of positive charge. For electrons, the magnitude is the same, but the sign is opposite, $I_{\text{electrons}} = -0.300 \times 10^{-3} \text{ C/s}$. Since each electron (e^-) has a charge of $-1.60 \times 10^{-19} \text{ C}$, we can convert the current in coulombs per second to electrons per second.

Solution

Starting with the definition of current, we have

Equation:

$$I_{\text{electrons}} = \frac{\Delta Q_{\text{electrons}}}{\Delta t} = \frac{-0.300 \times 10^{-3} \text{ C}}{\text{s}}.$$

We divide this by the charge per electron, so that

Equation:

$$\begin{aligned} \frac{e^-}{\text{s}} &= \frac{-0.300 \times 10^{-3} \text{ C}}{\text{s}} \times \frac{1 e^-}{-1.60 \times 10^{-19} \text{ C}} \\ &= 1.88 \times 10^{15} \frac{e^-}{\text{s}}. \end{aligned}$$

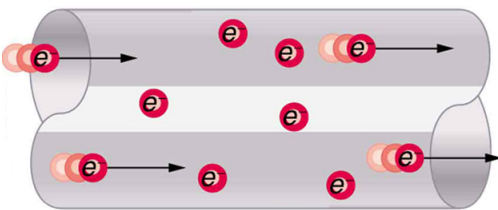
Discussion

There are so many charged particles moving, even in small currents, that individual charges are not noticed, just as individual water molecules are not noticed in water flow. Even more amazing is that they do not always keep moving forward like soldiers in a parade. Rather they are like a crowd of people with movement in different directions but a general trend to move forward. There are lots of collisions with atoms in the metal wire and, of course, with other electrons.

Drift Velocity

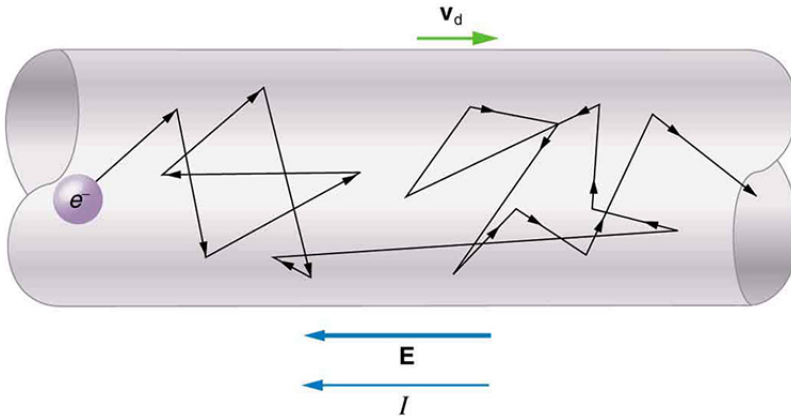
Electrical signals are known to move very rapidly. Telephone conversations carried by currents in wires cover large distances without noticeable delays. Lights come on as soon as a switch is flicked. Most electrical signals carried by currents travel at speeds on the order of 10^8 m/s, a significant fraction of the speed of light. Interestingly, the individual charges that make up the current move *much* more slowly on average, typically drifting at speeds on the order of 10^{-4} m/s. How do we reconcile these two speeds, and what does it tell us about standard conductors?

The high speed of electrical signals results from the fact that the force between charges acts rapidly at a distance. Thus, when a free charge is forced into a wire, as in [\[link\]](#), the incoming charge pushes other charges ahead of it, which in turn push on charges farther down the line. The density of charge in a system cannot easily be increased, and so the signal is passed on rapidly. The resulting electrical shock wave moves through the system at nearly the speed of light. To be precise, this rapidly moving signal or shock wave is a rapidly propagating change in electric field.



When charged particles are forced into this volume of a conductor, an equal number are quickly forced to leave. The repulsion between like charges makes it difficult to increase the number of charges in a volume. Thus, as one charge enters, another leaves almost immediately, carrying the signal rapidly forward.

Good conductors have large numbers of free charges in them. In metals, the free charges are free electrons. [\[link\]](#) shows how free electrons move through an ordinary conductor. The distance that an individual electron can move between collisions with atoms or other electrons is quite small. The electron paths thus appear nearly random, like the motion of atoms in a gas. But there is an electric field in the conductor that causes the electrons to drift in the direction shown (opposite to the field, since they are negative). The **drift velocity** v_d is the average velocity of the free charges. Drift velocity is quite small, since there are so many free charges. If we have an estimate of the density of free electrons in a conductor, we can calculate the drift velocity for a given current. The larger the density, the lower the velocity required for a given current.



Free electrons moving in a conductor make many collisions with other electrons and atoms. The path of one electron is shown. The average velocity of the free charges is called the drift velocity, v_d , and it is in the direction opposite to the electric field for electrons. The collisions normally transfer energy to the conductor, requiring a constant supply of energy to maintain a steady current.

Note:

Conduction of Electricity and Heat

Good electrical conductors are often good heat conductors, too. This is because large numbers of free electrons can carry electrical current and can transport thermal energy.

The free-electron collisions transfer energy to the atoms of the conductor. The electric field does work in moving the electrons through a distance, but that work does not increase the kinetic energy (nor speed, therefore) of the electrons. The work is transferred to the conductor's atoms, possibly

increasing temperature. Thus a continuous power input is required to keep a current flowing. An exception, of course, is found in superconductors, for reasons we shall explore in a later chapter. Superconductors can have a steady current without a continual supply of energy—a great energy savings. In contrast, the supply of energy can be useful, such as in a lightbulb filament. The supply of energy is necessary to increase the temperature of the tungsten filament, so that the filament glows.

Note:

Making Connections: Take-Home Investigation—Filament Observations

Find a lightbulb with a filament. Look carefully at the filament and describe its structure. To what points is the filament connected?

We can obtain an expression for the relationship between current and drift velocity by considering the number of free charges in a segment of wire, as illustrated in [\[link\]](#). The number of free charges per unit volume is given the symbol n and depends on the material. The shaded segment has a volume Ax , so that the number of free charges in it is nAx . The charge ΔQ in this segment is thus $qnAx$, where q is the amount of charge on each carrier. (Recall that for electrons, q is -1.60×10^{-19} C.) Current is charge moved per unit time; thus, if all the original charges move out of this segment in time Δt , the current is

Equation:

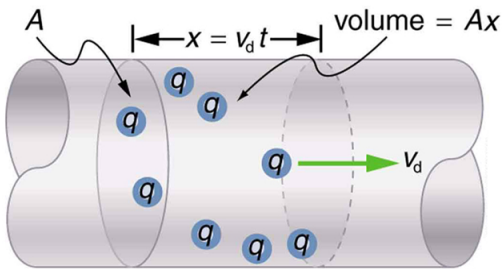
$$I = \frac{\Delta Q}{\Delta t} = \frac{qnAx}{\Delta t}.$$

Note that $x/\Delta t$ is the magnitude of the drift velocity, v_d , since the charges move an average distance x in a time Δt . Rearranging terms gives

Equation:

$$I = nqAv_d,$$

where I is the current through a wire of cross-sectional area A made of a material with a free charge density n . The carriers of the current each have charge q and move with a drift velocity of magnitude v_d .



All the charges in the shaded volume of this wire move out in a time t , having a drift velocity of magnitude $v_d = x/t$. See text for further discussion.

Note that simple drift velocity is not the entire story. The speed of an electron is much greater than its drift velocity. In addition, not all of the electrons in a conductor can move freely, and those that do might move somewhat faster or slower than the drift velocity. So what do we mean by free electrons? Atoms in a metallic conductor are packed in the form of a lattice structure. Some electrons are far enough away from the atomic nuclei that they do not experience the attraction of the nuclei as much as the inner electrons do. These are the free electrons. They are not bound to a single atom but can instead move freely among the atoms in a “sea” of electrons. These free electrons respond by accelerating when an electric field is applied. Of course as they move they collide with the atoms in the lattice and other electrons, generating thermal energy, and the conductor gets warmer. In an insulator, the organization of the atoms and the structure do not allow for such free electrons.

Example:**Calculating Drift Velocity in a Common Wire**

Calculate the drift velocity of electrons in a 12-gauge copper wire (which has a diameter of 2.053 mm) carrying a 20.0-A current, given that there is one free electron per copper atom. (Household wiring often contains 12-gauge copper wire, and the maximum current allowed in such wire is usually 20 A.) The density of copper is $8.80 \times 10^3 \text{ kg/m}^3$.

Strategy

We can calculate the drift velocity using the equation $I = nqAv_d$. The current $I = 20.0 \text{ A}$ is given, and $q = -1.60 \times 10^{-19} \text{ C}$ is the charge of an electron. We can calculate the area of a cross-section of the wire using the formula $A = \pi r^2$, where r is one-half the given diameter, 2.053 mm. We are given the density of copper, $8.80 \times 10^3 \text{ kg/m}^3$, and the periodic table shows that the atomic mass of copper is 63.54 g/mol. We can use these two quantities along with Avogadro's number, $6.02 \times 10^{23} \text{ atoms/mol}$, to determine n , the number of free electrons per cubic meter.

Solution

First, calculate the density of free electrons in copper. There is one free electron per copper atom. Therefore, is the same as the number of copper atoms per m^3 . We can now find n as follows:

Equation:

$$\begin{aligned} n &= \frac{1 e^-}{\text{atom}} \times \frac{6.02 \times 10^{23} \text{ atoms}}{\text{mol}} \times \frac{1 \text{ mol}}{63.54 \text{ g}} \times \frac{1000 \text{ g}}{\text{kg}} \times \frac{8.80 \times 10^3 \text{ kg}}{1 \text{ m}^3} \\ &= 8.342 \times 10^{28} e^-/\text{m}^3. \end{aligned}$$

The cross-sectional area of the wire is

Equation:

$$\begin{aligned} A &= \pi r^2 \\ &= \pi \left(\frac{2.053 \times 10^{-3} \text{ m}}{2} \right)^2 \\ &= 3.310 \times 10^{-6} \text{ m}^2. \end{aligned}$$

Rearranging $I = nqAv_d$ to isolate drift velocity gives

Equation:

$$\begin{aligned}
 v_d &= \frac{I}{nqA} \\
 &= \frac{20.0 \text{ A}}{(8.342 \times 10^{28} / \text{m}^3)(-1.60 \times 10^{-19} \text{ C})(3.310 \times 10^{-6} \text{ m}^2)} \\
 &= -4.53 \times 10^{-4} \text{ m/s}.
 \end{aligned}$$

Discussion

The minus sign indicates that the negative charges are moving in the direction opposite to conventional current. The small value for drift velocity (on the order of 10^{-4} m/s) confirms that the signal moves on the order of 10^{12} times faster (about 10^8 m/s) than the charges that carry it.

Section Summary

- Electric current I is the rate at which charge flows, given by

Equation:

$$I = \frac{\Delta Q}{\Delta t},$$

where ΔQ is the amount of charge passing through an area in time Δt .

- The direction of conventional current is taken as the direction in which positive charge moves.
- The SI unit for current is the ampere (A), where $1 \text{ A} = 1 \text{ C/s}$.
- Current is the flow of free charges, such as electrons and ions.
- Drift velocity v_d is the average speed at which these charges move.
- Current I is proportional to drift velocity v_d , as expressed in the relationship $I = nqAv_d$. Here, I is the current through a wire of cross-sectional area A . The wire's material has a free-charge density n , and each carrier has charge q and a drift velocity v_d .
- Electrical signals travel at speeds about 10^{12} times greater than the drift velocity of free electrons.

Conceptual Questions

Exercise:

Problem:

Can a wire carry a current and still be neutral—that is, have a total charge of zero? Explain.

Exercise:**Problem:**

Car batteries are rated in ampere-hours ($A \cdot h$). To what physical quantity do ampere-hours correspond (voltage, charge, . . .), and what relationship do ampere-hours have to energy content?

Exercise:**Problem:**

If two different wires having identical cross-sectional areas carry the same current, will the drift velocity be higher or lower in the better conductor? Explain in terms of the equation $v_d = \frac{I}{nqA}$, by considering how the density of charge carriers n relates to whether or not a material is a good conductor.

Exercise:**Problem:**

Why are two conducting paths from a voltage source to an electrical device needed to operate the device?

Exercise:**Problem:**

In cars, one battery terminal is connected to the metal body. How does this allow a single wire to supply current to electrical devices rather than two wires?

Exercise:

Problem:

Why isn't a bird sitting on a high-voltage power line electrocuted? Contrast this with the situation in which a large bird hits two wires simultaneously with its wings.

Problems & Exercises**Exercise:****Problem:**

What is the current in milliamperes produced by the solar cells of a pocket calculator through which 4.00 C of charge passes in 4.00 h?

Solution:

0.278 mA

Exercise:**Problem:**

A total of 600 C of charge passes through a flashlight in 0.500 h. What is the average current?

Exercise:**Problem:**

What is the current when a typical static charge of $0.250\ \mu\text{C}$ moves from your finger to a metal doorknob in $1.00\ \mu\text{s}$?

Solution:

0.250 A

Exercise:

Problem:

Find the current when 2.00 nC jumps between your comb and hair over a 0.500 - μ s time interval.

Exercise:**Problem:**

A large lightning bolt had a 20,000-A current and moved 30.0 C of charge. What was its duration?

Solution:

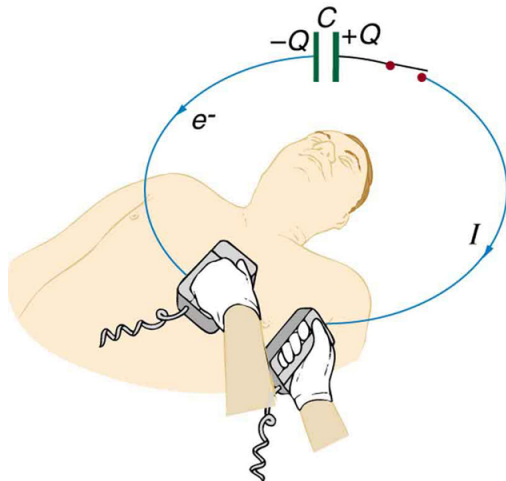
1.50ms

Exercise:**Problem:**

The 200-A current through a spark plug moves 0.300 mC of charge. How long does the spark last?

Exercise:**Problem:**

(a) A defibrillator sends a 6.00-A current through the chest of a patient by applying a 10,000-V potential as in the figure below. What is the resistance of the path? (b) The defibrillator paddles make contact with the patient through a conducting gel that greatly reduces the path resistance. Discuss the difficulties that would ensue if a larger voltage were used to produce the same current through the patient, but with the path having perhaps 50 times the resistance. (Hint: The current must be about the same, so a higher voltage would imply greater power. Use this equation for power: $P = I^2 R$.)



The capacitor in a defibrillation unit drives a current through the heart of a patient.

Solution:

(a) $1.67\text{k}\Omega$

(b) If a 50 times larger resistance existed, keeping the current about the same, the power would be increased by a factor of about 50 (based on the equation $P = I^2 R$), causing much more energy to be transferred to the skin, which could cause serious burns. The gel used reduces the resistance, and therefore reduces the power transferred to the skin.

Exercise:

Problem:

During open-heart surgery, a defibrillator can be used to bring a patient out of cardiac arrest. The resistance of the path is $500\ \Omega$ and a 10.0-mA current is needed. What voltage should be applied?

Exercise:

Problem:

(a) A defibrillator passes 12.0 A of current through the torso of a person for 0.0100 s. How much charge moves? (b) How many electrons pass through the wires connected to the patient? (See figure two problems earlier.)

Solution:

(a) 0.120 C

(b) 7.50×10^{17} electrons

Exercise:**Problem:**

A clock battery wears out after moving 10,000 C of charge through the clock at a rate of 0.500 mA. (a) How long did the clock run? (b) How many electrons per second flowed?

Exercise:**Problem:**

The batteries of a submerged non-nuclear submarine supply 1000 A at full speed ahead. How long does it take to move Avogadro's number (6.02×10^{23}) of electrons at this rate?

Solution:

96.3 s

Exercise:

Problem:

Electron guns are used in X-ray tubes. The electrons are accelerated through a relatively large voltage and directed onto a metal target, producing X-rays. (a) How many electrons per second strike the target if the current is 0.500 mA? (b) What charge strikes the target in 0.750 s?

Exercise:**Problem:**

A large cyclotron directs a beam of He^{++} nuclei onto a target with a beam current of 0.250 mA. (a) How many He^{++} nuclei per second is this? (b) How long does it take for 1.00 C to strike the target? (c) How long before 1.00 mol of He^{++} nuclei strike the target?

Solution:

(a) $7.81 \times 10^{14} \text{ He}^{++} \text{ nuclei/s}$

(b) $4.00 \times 10^3 \text{ s}$

(c) $7.71 \times 10^8 \text{ s}$

Exercise:**Problem:**

Repeat the above example on [\[link\]](#), but for a wire made of silver and given there is one free electron per silver atom.

Exercise:**Problem:**

Using the results of the above example on [\[link\]](#), find the drift velocity in a copper wire of twice the diameter and carrying 20.0 A.

Solution:

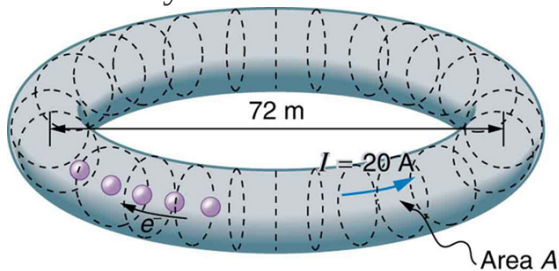
$$-1.13 \times 10^{-4} \text{ m/s}$$

Exercise:**Problem:**

A 14-gauge copper wire has a diameter of 1.628 mm. What magnitude current flows when the drift velocity is 1.00 mm/s? (See above example on [\[link\]](#) for useful information.)

Exercise:**Problem:**

SPEAR, a storage ring about 72.0 m in diameter at the Stanford Linear Accelerator (closed in 2009), has a 20.0-A circulating beam of electrons that are moving at nearly the speed of light. (See [\[link\]](#).) How many electrons are in the beam?



Electrons circulating in the storage ring called SPEAR constitute a 20.0-A current. Because they travel close to the speed of light, each electron completes many orbits in each second.

Solution:

$$9.42 \times 10^{13} \text{ electrons}$$

Glossary

electric current

the rate at which charge flows, $I = \Delta Q / \Delta t$

ampere

(amp) the SI unit for current; $1 \text{ A} = 1 \text{ C/s}$

drift velocity

the average velocity at which free charges flow in response to an electric field

Ohm's Law: Resistance and Simple Circuits

- Explain the origin of Ohm's law.
- Calculate voltages, currents, or resistances with Ohm's law.
- Explain what an ohmic material is.
- Describe a simple circuit.

What drives current? We can think of various devices—such as batteries, generators, wall outlets, and so on—which are necessary to maintain a current. All such devices create a potential difference and are loosely referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference V that creates an electric field. The electric field in turn exerts force on charges, causing current.

Ohm's Law

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is *directly proportional to the voltage applied*:

Equation:

$$I \propto V.$$

This important relationship is known as **Ohm's law**. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law like that for friction—an experimentally observed phenomenon. Such a linear relationship doesn't always occur.

Resistance and Simple Circuits

If voltage drives current, what impedes it? The electric property that impedes current (crudely similar to friction and air resistance) is called **resistance** R . Collisions of moving charges with atoms and molecules in a substance transfer energy to the substance and limit current. Resistance is defined as inversely proportional to current, or

Equation:

$$I \propto \frac{1}{R}.$$

Thus, for example, current is cut in half if resistance doubles. Combining the relationships of current to voltage and current to resistance gives

Equation:

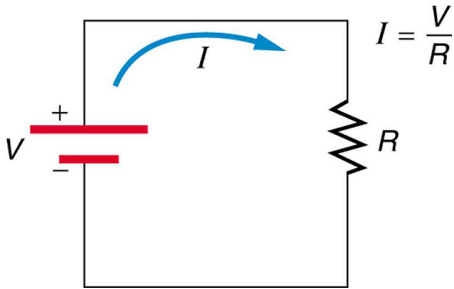
$$I = \frac{V}{R}.$$

This relationship is also called Ohm's law. Ohm's law in this form really defines resistance for certain materials. Ohm's law (like Hooke's law) is not universally valid. The many substances for which Ohm's law holds are called **ohmic**. These include good conductors like copper and aluminum, and some poor conductors under certain circumstances. Ohmic materials have a resistance R that is independent of voltage V and current I . An object that has simple resistance is called a *resistor*, even if its resistance is small. The unit for resistance is an **ohm** and is given the symbol Ω (upper case Greek omega). Rearranging $I = V/R$ gives $R = V/I$, and so the units of resistance are 1 ohm = 1 volt per ampere:

Equation:

$$1 \Omega = 1 \frac{V}{A}.$$

[\[link\]](#) shows the schematic for a simple circuit. A **simple circuit** has a single voltage source and a single resistor. The wires connecting the voltage source to the resistor can be assumed to have negligible resistance, or their resistance can be included in R .



A simple electric circuit in which a closed path for current to flow is supplied by conductors (usually metal wires) connecting a load to the terminals of a battery, represented by the red parallel lines.

The zigzag symbol represents the single resistor and includes any resistance in the connections to the voltage source.

Example:**Calculating Resistance: An Automobile Headlight**

What is the resistance of an automobile headlight through which 2.50 A flows when 12.0 V is applied to it?

Strategy

We can rearrange Ohm's law as stated by $I = V/R$ and use it to find the resistance.

Solution

Rearranging $I = V/R$ and substituting known values gives

Equation:

$$R = \frac{V}{I} = \frac{12.0 \text{ V}}{2.50 \text{ A}} = 4.80 \Omega.$$

Discussion

This is a relatively small resistance, but it is larger than the cold resistance of the headlight. As we shall see in [Resistance and Resistivity](#), resistance usually increases with temperature, and so the bulb has a lower resistance when it is first switched on and will draw considerably more current during its brief warm-up period.

Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of $10^{12} \Omega$ or more. A dry person may have a hand-to-foot resistance of $10^5 \Omega$, whereas the resistance of the human heart is about $10^3 \Omega$. A meter-long piece of large-diameter copper wire may have a resistance of $10^{-5} \Omega$, and superconductors have no resistance at all (they are non-ohmic). Resistance is related to the shape of an object and the material of which it is composed, as will be seen in [Resistance and Resistivity](#).

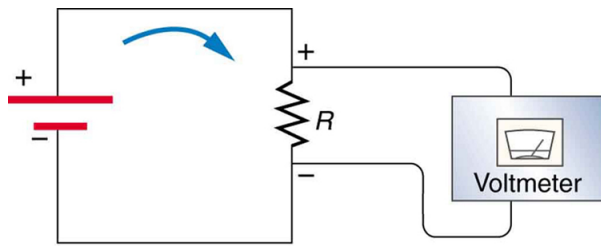
Additional insight is gained by solving $I = V/R$ for V , yielding

Equation:

$$V = IR.$$

This expression for V can be interpreted as the *voltage drop across a resistor produced by the flow of current I* . The phrase *IR drop* is often used for this voltage. For instance, the headlight in [\[link\]](#) has an IR drop of 12.0 V. If voltage is measured at various points in a circuit, it will be seen to increase at the voltage source and decrease at the resistor. Voltage is similar to fluid pressure. The voltage source is like a pump, creating a pressure difference, causing current—the flow of charge. The resistor is like a pipe that reduces pressure and limits flow because of its resistance. Conservation of energy has important consequences here. The voltage source supplies

energy (causing an electric field and a current), and the resistor converts it to another form (such as thermal energy). In a simple circuit (one with a single simple resistor), the voltage supplied by the source equals the voltage drop across the resistor, since $PE = q\Delta V$, and the same q flows through each. Thus the energy supplied by the voltage source and the energy converted by the resistor are equal. (See [\[link\]](#).)



$$V = IR = 18 \text{ V}$$

The voltage drop across a resistor in a simple circuit equals the voltage output of the battery.

Note:**Making Connections: Conservation of Energy**

In a simple electrical circuit, the sole resistor converts energy supplied by the source into another form. Conservation of energy is evidenced here by the fact that all of the energy supplied by the source is converted to another form by the resistor alone. We will find that conservation of energy has other important applications in circuits and is a powerful tool in circuit analysis.

Note:

PhET Explorations: Ohm's Law

See how the equation form of Ohm's law relates to a simple circuit. Adjust the voltage and resistance, and see the current change according to Ohm's law. The sizes of the symbols in the equation change to match the circuit diagram.

https://phet.colorado.edu/sims/html/ohms-law/latest/ohms-law_en.html

Section Summary

- A simple circuit *is* one in which there is a single voltage source and a single resistance.
- One statement of Ohm's law gives the relationship between current I , voltage V , and resistance R in a simple circuit to be $I = \frac{V}{R}$.
- Resistance has units of ohms (Ω), related to volts and amperes by $1 \Omega = 1 \text{ V/A}$.
- There is a voltage or IR drop across a resistor, caused by the current flowing through it, given by $V = IR$.

Conceptual Questions

Exercise:

Problem:

The IR drop across a resistor means that there is a change in potential or voltage across the resistor. Is there any change in current as it passes through a resistor? Explain.

Exercise:

Problem:

How is the IR drop in a resistor similar to the pressure drop in a fluid flowing through a pipe?

Problems & Exercises

Exercise:**Problem:**

What current flows through the bulb of a 3.00-V flashlight when its hot resistance is $3.60\ \Omega$?

Solution:

0.833 A

Exercise:**Problem:**

Calculate the effective resistance of a pocket calculator that has a 1.35-V battery and through which 0.200 mA flows.

Exercise:**Problem:**

What is the effective resistance of a car's starter motor when 150 A flows through it as the car battery applies 11.0 V to the motor?

Solution:

$7.33 \times 10^{-2}\ \Omega$

Exercise:**Problem:**

How many volts are supplied to operate an indicator light on a DVD player that has a resistance of $140\ \Omega$, given that 25.0 mA passes through it?

Exercise:

Problem:

(a) Find the voltage drop in an extension cord having a $0.0600\text{-}\Omega$ resistance and through which 5.00 A is flowing. (b) A cheaper cord utilizes thinner wire and has a resistance of $0.300\text{ }\Omega$. What is the voltage drop in it when 5.00 A flows? (c) Why is the voltage to whatever appliance is being used reduced by this amount? What is the effect on the appliance?

Solution:

(a) 0.300 V

(b) 1.50 V

(c) The voltage supplied to whatever appliance is being used is reduced because the total voltage drop from the wall to the final output of the appliance is fixed. Thus, if the voltage drop across the extension cord is large, the voltage drop across the appliance is significantly decreased, so the power output by the appliance can be significantly decreased, reducing the ability of the appliance to work properly.

Exercise:**Problem:**

A power transmission line is hung from metal towers with glass insulators having a resistance of $1.00 \times 10^9\text{ }\Omega$. What current flows through the insulator if the voltage is 200 kV ? (Some high-voltage lines are DC.)

Glossary**Ohm's law**

an empirical relation stating that the current I is proportional to the potential difference V , $\propto V$; it is often written as $I = V/R$, where R is the resistance

resistance

the electric property that impedes current; for ohmic materials, it is the ratio of voltage to current, $R = V/I$

ohm

the unit of resistance, given by $1\Omega = 1 \text{ V/A}$

ohmic

a type of a material for which Ohm's law is valid

simple circuit

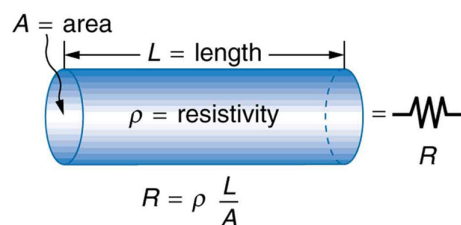
a circuit with a single voltage source and a single resistor

Resistance and Resistivity

- Explain the concept of resistivity.
- Use resistivity to calculate the resistance of specified configurations of material.
- Use the thermal coefficient of resistivity to calculate the change of resistance with temperature.

Material and Shape Dependence of Resistance

The resistance of an object depends on its shape and the material of which it is composed. The cylindrical resistor in [\[link\]](#) is easy to analyze, and, by so doing, we can gain insight into the resistance of more complicated shapes. As you might expect, the cylinder's electric resistance R is directly proportional to its length L , similar to the resistance of a pipe to fluid flow. The longer the cylinder, the more collisions charges will make with its atoms. The greater the diameter of the cylinder, the more current it can carry (again similar to the flow of fluid through a pipe). In fact, R is inversely proportional to the cylinder's cross-sectional area A .



A uniform cylinder of length L and cross-sectional area A . Its resistance to the flow of current is similar to the resistance posed by a pipe to fluid flow.

The longer the cylinder, the greater its

resistance. The larger its cross-sectional area A , the smaller its resistance.

For a given shape, the resistance depends on the material of which the object is composed. Different materials offer different resistance to the flow of charge. We define the **resistivity** ρ of a substance so that the **resistance** R of an object is directly proportional to ρ . Resistivity ρ is an *intrinsic* property of a material, independent of its shape or size. The resistance R of a uniform cylinder of length L , of cross-sectional area A , and made of a material with resistivity ρ , is

Equation:

$$R = \frac{\rho L}{A}.$$

[\[link\]](#) gives representative values of ρ . The materials listed in the table are separated into categories of conductors, semiconductors, and insulators, based on broad groupings of resistivities. Conductors have the smallest resistivities, and insulators have the largest; semiconductors have intermediate resistivities. Conductors have varying but large free charge densities, whereas most charges in insulators are bound to atoms and are not free to move. Semiconductors are intermediate, having far fewer free charges than conductors, but having properties that make the number of free charges depend strongly on the type and amount of impurities in the semiconductor. These unique properties of semiconductors are put to use in modern electronics, as will be explored in later chapters.

Material	Resistivity ρ ($\Omega \cdot \text{m}$)
<i>Conductors</i>	
Silver	1.59×10^{-8}
Copper	1.72×10^{-8}
Gold	2.44×10^{-8}
Aluminum	2.65×10^{-8}
Tungsten	5.6×10^{-8}
Iron	9.71×10^{-8}
Platinum	10.6×10^{-8}
Steel	20×10^{-8}
Lead	22×10^{-8}

Material	Resistivity ρ ($\Omega \cdot \text{m}$)
Manganin (Cu, Mn, Ni alloy)	44×10^{-8}
Constantan (Cu, Ni alloy)	49×10^{-8}
Mercury	96×10^{-8}
Nichrome (Ni, Fe, Cr alloy)	100×10^{-8}
<i>Semiconductors</i> [footnote] Values depend strongly on amounts and types of impurities	
Carbon (pure)	3.5×10^{-5}
Carbon	$(3.5 - 60) \times 10^{-5}$
Germanium (pure)	600×10^{-3}
Germanium	$(1 - 600) \times 10^{-3}$

Material	Resistivity ρ ($\Omega \cdot \text{m}$)
Silicon (pure)	2300
Silicon	0.1–2300
<i>Insulators</i>	
Amber	5×10^{14}
Glass	$10^9 - 10^{14}$
Lucite	$>10^{13}$
Mica	$10^{11} - 10^{15}$
Quartz (fused)	75×10^{16}
Rubber (hard)	$10^{13} - 10^{16}$
Sulfur	10^{15}

Material	Resistivity ρ ($\Omega \cdot \text{m}$)
Teflon	$>10^{13}$
Wood	$10^8 - 10^{11}$

Resistivities ρ of Various materials at 20°C

Example:

Calculating Resistor Diameter: A Headlight Filament

A car headlight filament is made of tungsten and has a cold resistance of 0.350Ω . If the filament is a cylinder 4.00 cm long (it may be coiled to save space), what is its diameter?

Strategy

We can rearrange the equation $R = \frac{\rho L}{A}$ to find the cross-sectional area A of the filament from the given information. Then its diameter can be found by assuming it has a circular cross-section.

Solution

The cross-sectional area, found by rearranging the expression for the resistance of a cylinder given in $R = \frac{\rho L}{A}$, is

Equation:

$$A = \frac{\rho L}{R}.$$

Substituting the given values, and taking ρ from [\[link\]](#), yields

Equation:

$$\begin{aligned} A &= \frac{(5.6 \times 10^{-8} \Omega \cdot \text{m})(4.00 \times 10^{-2} \text{ m})}{0.350 \Omega} \\ &= 6.40 \times 10^{-9} \text{ m}^2. \end{aligned}$$

The area of a circle is related to its diameter D by

Equation:

$$A = \frac{\pi D^2}{4}.$$

Solving for the diameter D , and substituting the value found for A , gives

Equation:

$$\begin{aligned} D &= 2\left(\frac{A}{\pi}\right)^{\frac{1}{2}} = 2\left(\frac{6.40 \times 10^{-9} \text{ m}^2}{3.14}\right)^{\frac{1}{2}} \\ &= 9.0 \times 10^{-5} \text{ m}. \end{aligned}$$

Discussion

The diameter is just under a tenth of a millimeter. It is quoted to only two digits, because ρ is known to only two digits.

Temperature Variation of Resistance

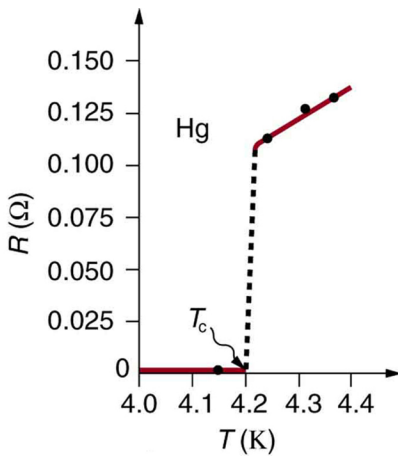
The resistivity of all materials depends on temperature. Some even become superconductors (zero resistivity) at very low temperatures. (See [\[link\]](#).) Conversely, the resistivity of conductors increases with increasing temperature. Since the atoms vibrate more rapidly and over larger distances at higher temperatures, the electrons moving through a metal make more collisions, effectively making the resistivity higher. Over relatively small temperature changes (about 100°C or less), resistivity ρ varies with temperature change ΔT as expressed in the following equation

Equation:

$$\rho = \rho_0(1 + \alpha\Delta T),$$

where ρ_0 is the original resistivity and α is the **temperature coefficient of resistivity**. (See the values of α in [\[link\]](#) below.) For larger temperature changes, α may vary or a nonlinear equation may be needed to find ρ . Note

that α is positive for metals, meaning their resistivity increases with temperature. Some alloys have been developed specifically to have a small temperature dependence. Manganin (which is made of copper, manganese and nickel), for example, has α close to zero (to three digits on the scale in [\[link\]](#)), and so its resistivity varies only slightly with temperature. This is useful for making a temperature-independent resistance standard, for example.



The resistance of a sample of mercury is zero at very low temperatures—it is a superconductor up to about 4.2 K. Above that critical temperature, its resistance makes a sudden jump and then increases nearly linearly with temperature.

Material	Coefficient $\alpha(1/^{\circ}\text{C})$ [footnote] Values at 20°C.
<i>Conductors</i>	
Silver	3.8×10^{-3}
Copper	3.9×10^{-3}
Gold	3.4×10^{-3}
Aluminum	3.9×10^{-3}
Tungsten	4.5×10^{-3}
Iron	5.0×10^{-3}
Platinum	3.93×10^{-3}
Lead	3.9×10^{-3}
Manganin (Cu, Mn, Ni alloy)	0.000×10^{-3}

Material	Coefficient $\alpha(1/^{\circ}\text{C})$ [footnote] Values at 20°C.
Constantan (Cu, Ni alloy)	0.002×10^{-3}
Mercury	0.89×10^{-3}
Nichrome (Ni, Fe, Cr alloy)	0.4×10^{-3}
<i>Semiconductors</i>	
Carbon (pure)	-0.5×10^{-3}
Germanium (pure)	-50×10^{-3}
Silicon (pure)	-70×10^{-3}

Temperature Coefficients of Resistivity α

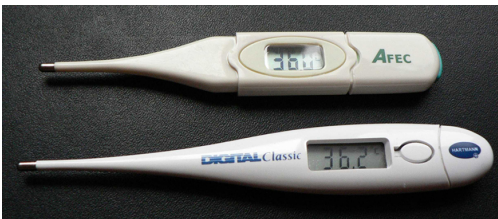
Note also that α is negative for the semiconductors listed in [\[link\]](#), meaning that their resistivity decreases with increasing temperature. They become better conductors at higher temperature, because increased thermal agitation increases the number of free charges available to carry current. This property of decreasing ρ with temperature is also related to the type and amount of impurities present in the semiconductors.

The resistance of an object also depends on temperature, since R_0 is directly proportional to ρ . For a cylinder we know $R = \rho L/A$, and so, if L and A do not change greatly with temperature, R will have the same temperature dependence as ρ . (Examination of the coefficients of linear expansion shows them to be about two orders of magnitude less than typical temperature coefficients of resistivity, and so the effect of temperature on L and A is about two orders of magnitude less than on ρ .) Thus,

Equation:

$$R = R_0(1 + \alpha\Delta T)$$

is the temperature dependence of the resistance of an object, where R_0 is the original resistance and R is the resistance after a temperature change ΔT . Numerous thermometers are based on the effect of temperature on resistance. (See [\[link\]](#).) One of the most common is the thermistor, a semiconductor crystal with a strong temperature dependence, the resistance of which is measured to obtain its temperature. The device is small, so that it quickly comes into thermal equilibrium with the part of a person it touches.



These familiar
thermometers are based
on the automated
measurement of a
thermistor's temperature-
dependent resistance.
(credit: Biol, Wikimedia
Commons)

Example:**Calculating Resistance: Hot-Filament Resistance**

Although caution must be used in applying $\rho = \rho_0(1 + \alpha\Delta T)$ and $R = R_0(1 + \alpha\Delta T)$ for temperature changes greater than 100°C, for tungsten the equations work reasonably well for very large temperature changes. What, then, is the resistance of the tungsten filament in the previous example if its temperature is increased from room temperature (20°C) to a typical operating temperature of 2850°C?

Strategy

This is a straightforward application of $R = R_0(1 + \alpha\Delta T)$, since the original resistance of the filament was given to be $R_0 = 0.350 \, \Omega$, and the temperature change is $\Delta T = 2830^\circ\text{C}$.

Solution

The hot resistance R is obtained by entering known values into the above equation:

Equation:

$$\begin{aligned} R &= R_0(1 + \alpha\Delta T) \\ &= (0.350 \, \Omega)[1 + (4.5 \times 10^{-3}/^\circ\text{C})(2830^\circ\text{C})] \\ &= 4.8 \, \Omega. \end{aligned}$$

Discussion

This value is consistent with the headlight resistance example in [Ohm's Law: Resistance and Simple Circuits](#).

Note:**PhET Explorations: Resistance in a Wire**

Learn about the physics of resistance in a wire. Change its resistivity, length, and area to see how they affect the wire's resistance. The sizes of the symbols in the equation change along with the diagram of a wire.

Section Summary

- The resistance R of a cylinder of length L and cross-sectional area A is $R = \frac{\rho L}{A}$, where ρ is the resistivity of the material.
- Values of ρ in [\[link\]](#) show that materials fall into three groups—*conductors, semiconductors, and insulators*.
- Temperature affects resistivity; for relatively small temperature changes ΔT , resistivity is $\rho = \rho_0(1 + \alpha\Delta T)$, where ρ_0 is the original resistivity and α is the temperature coefficient of resistivity.
- [\[link\]](#) gives values for α , the temperature coefficient of resistivity.
- The resistance R of an object also varies with temperature: $R = R_0(1 + \alpha\Delta T)$, where R_0 is the original resistance, and R is the resistance after the temperature change.

Conceptual Questions

Exercise:

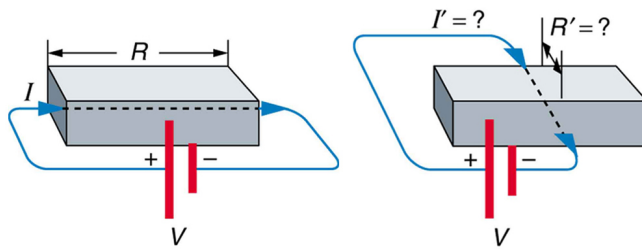
Problem:

In which of the three semiconducting materials listed in [\[link\]](#) do impurities supply free charges? (Hint: Examine the range of resistivity for each and determine whether the pure semiconductor has the higher or lower conductivity.)

Exercise:

Problem:

Does the resistance of an object depend on the path current takes through it? Consider, for example, a rectangular bar—is its resistance the same along its length as across its width? (See [\[link\]](#).)



Does current taking two different paths through the same object encounter different resistance?

Exercise:

Problem:

If aluminum and copper wires of the same length have the same resistance, which has the larger diameter? Why?

Exercise:

Problem:

Explain why $R = R_0(1 + \alpha\Delta T)$ for the temperature variation of the resistance R of an object is not as accurate as $\rho = \rho_0(1 + \alpha\Delta T)$, which gives the temperature variation of resistivity ρ .

Problems & Exercises

Exercise:

Problem:

What is the resistance of a 20.0-m-long piece of 12-gauge copper wire having a 2.053-mm diameter?

Solution:

0.104 Ω

Exercise:**Problem:**

The diameter of 0-gauge copper wire is 8.252 mm. Find the resistance of a 1.00-km length of such wire used for power transmission.

Exercise:**Problem:**

If the 0.100-mm diameter tungsten filament in a light bulb is to have a resistance of $0.200\ \Omega$ at 20.0°C , how long should it be?

Solution:

$$2.8 \times 10^{-2}\ \text{m}$$

Exercise:**Problem:**

Find the ratio of the diameter of aluminum to copper wire, if they have the same resistance per unit length (as they might in household wiring).

Exercise:**Problem:**

What current flows through a 2.54-cm-diameter rod of pure silicon that is 20.0 cm long, when $1.00 \times 10^3\ \text{V}$ is applied to it? (Such a rod may be used to make nuclear-particle detectors, for example.)

Solution:

$$1.10 \times 10^{-3}\ \text{A}$$

Exercise:

Problem:

(a) To what temperature must you raise a copper wire, originally at 20.0°C , to double its resistance, neglecting any changes in dimensions? (b) Does this happen in household wiring under ordinary circumstances?

Exercise:**Problem:**

A resistor made of Nichrome wire is used in an application where its resistance cannot change more than 1.00% from its value at 20.0°C . Over what temperature range can it be used?

Solution:

-5°C to 45°C

Exercise:**Problem:**

Of what material is a resistor made if its resistance is 40.0% greater at 100°C than at 20.0°C ?

Exercise:**Problem:**

An electronic device designed to operate at any temperature in the range from -10.0°C to 55.0°C contains pure carbon resistors. By what factor does their resistance increase over this range?

Solution:

1.03

Exercise:

Problem:

(a) Of what material is a wire made, if it is 25.0 m long with a 0.100 mm diameter and has a resistance of $77.7\ \Omega$ at 20.0°C ? (b) What is its resistance at 150°C ?

Exercise:**Problem:**

Assuming a constant temperature coefficient of resistivity, what is the maximum percent decrease in the resistance of a constantan wire starting at 20.0°C ?

Solution:

0.06%

Exercise:**Problem:**

A wire is drawn through a die, stretching it to four times its original length. By what factor does its resistance increase?

Exercise:**Problem:**

A copper wire has a resistance of $0.500\ \Omega$ at 20.0°C , and an iron wire has a resistance of $0.525\ \Omega$ at the same temperature. At what temperature are their resistances equal?

Solution:

-17°C

Exercise:

Problem:

(a) Digital medical thermometers determine temperature by measuring the resistance of a semiconductor device called a thermistor (which has $\alpha = -0.0600/^{\circ}\text{C}$) when it is at the same temperature as the patient. What is a patient's temperature if the thermistor's resistance at that temperature is 82.0% of its value at 37.0°C (normal body temperature)? (b) The negative value for α may not be maintained for very low temperatures. Discuss why and whether this is the case here. (Hint: Resistance can't become negative.)

Exercise:**Problem: Integrated Concepts**

(a) Redo [\[link\]](#) taking into account the thermal expansion of the tungsten filament. You may assume a thermal expansion coefficient of $12 \times 10^{-6}/^{\circ}\text{C}$. (b) By what percentage does your answer differ from that in the example?

Solution:

(a) $4.7 \, \Omega$ (total)

(b) 3.0% decrease

Exercise:**Problem: Unreasonable Results**

(a) To what temperature must you raise a resistor made of constantan to double its resistance, assuming a constant temperature coefficient of resistivity? (b) To cut it in half? (c) What is unreasonable about these results? (d) Which assumptions are unreasonable, or which premises are inconsistent?

Glossary

resistivity

an intrinsic property of a material, independent of its shape or size,
directly proportional to the resistance, denoted by ρ

temperature coefficient of resistivity

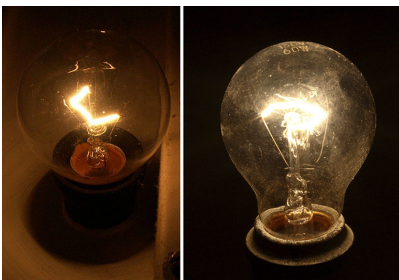
an empirical quantity, denoted by α , which describes the change in
resistance or resistivity of a material with temperature

Electric Power and Energy

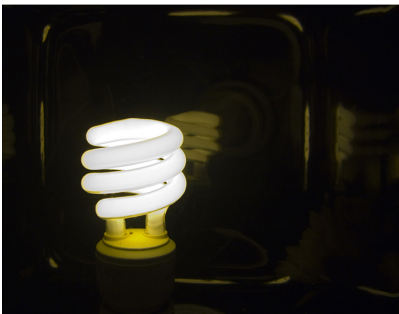
- Calculate the power dissipated by a resistor and power supplied by a power supply.
- Calculate the cost of electricity under various circumstances.

Power in Electric Circuits

Power is associated by many people with electricity. Knowing that power is the rate of energy use or energy conversion, what is the expression for **electric power**? Power transmission lines might come to mind. We also think of lightbulbs in terms of their power ratings in watts. Let us compare a 25-W bulb with a 60-W bulb. (See [\[link\]](#)(a).) Since both operate on the same voltage, the 60-W bulb must draw more current to have a greater power rating. Thus the 60-W bulb's resistance must be lower than that of a 25-W bulb. If we increase voltage, we also increase power. For example, when a 25-W bulb that is designed to operate on 120 V is connected to 240 V, it briefly glows very brightly and then burns out. Precisely how are voltage, current, and resistance related to electric power?



(a)



(b)

(a) Which of these lightbulbs, the 25-W bulb (upper left) or the 60-W bulb (upper right), has the higher resistance? Which draws more current? Which uses the most energy? Can you tell from the color that the 25-W filament is cooler? Is the brighter bulb a different color and if so why?

(credits: Dickbauch, Wikimedia Commons; Greg Westfall, Flickr) (b)

This compact fluorescent light (CFL) puts out the same intensity of light as the 60-W bulb, but at 1/4 to 1/10 the input power. (credit: dbgg1979, Flickr)

Electric energy depends on both the voltage involved and the charge moved. This is expressed most simply as $PE = qV$, where q is the charge moved and V is the voltage (or more precisely, the potential difference the

charge moves through). Power is the rate at which energy is moved, and so electric power is

Equation:

$$P = \frac{PE}{t} = \frac{qV}{t}.$$

Recognizing that current is $I = q/t$ (note that $\Delta t = t$ here), the expression for power becomes

Equation:

$$P = IV.$$

Electric power (P) is simply the product of current times voltage. Power has familiar units of watts. Since the SI unit for potential energy (PE) is the joule, power has units of joules per second, or watts. Thus, $1 \text{ A} \cdot \text{V} = 1 \text{ W}$. For example, cars often have one or more auxiliary power outlets with which you can charge a cell phone or other electronic devices. These outlets may be rated at 20 A, so that the circuit can deliver a maximum power $P = IV = (20 \text{ A})(12 \text{ V}) = 240 \text{ W}$. In some applications, electric power may be expressed as volt-amperes or even kilovolt-amperes ($1 \text{ kA} \cdot \text{V} = 1 \text{ kW}$).

To see the relationship of power to resistance, we combine Ohm's law with $P = IV$. Substituting $I = V/R$ gives $P = (V/R)V = V^2/R$. Similarly, substituting $V = IR$ gives $P = I(IR) = I^2R$. Three expressions for electric power are listed together here for convenience:

Equation:

$$P = IV$$

Equation:

$$P = \frac{V^2}{R}$$

Equation:

$$P = I^2 R.$$

Note that the first equation is always valid, whereas the other two can be used only for resistors. In a simple circuit, with one voltage source and a single resistor, the power supplied by the voltage source and that dissipated by the resistor are identical. (In more complicated circuits, P can be the power dissipated by a single device and not the total power in the circuit.)

Different insights can be gained from the three different expressions for electric power. For example, $P = V^2/R$ implies that the lower the resistance connected to a given voltage source, the greater the power delivered. Furthermore, since voltage is squared in $P = V^2/R$, the effect of applying a higher voltage is perhaps greater than expected. Thus, when the voltage is doubled to a 25-W bulb, its power nearly quadruples to about 100 W, burning it out. If the bulb's resistance remained constant, its power would be exactly 100 W, but at the higher temperature its resistance is higher, too.

Example:**Calculating Power Dissipation and Current: Hot and Cold Power**

(a) Consider the examples given in [Ohm's Law: Resistance and Simple Circuits](#) and [Resistance and Resistivity](#). Then find the power dissipated by the car headlight in these examples, both when it is hot and when it is cold.
(b) What current does it draw when cold?

Strategy for (a)

For the hot headlight, we know voltage and current, so we can use $P = IV$ to find the power. For the cold headlight, we know the voltage and resistance, so we can use $P = V^2/R$ to find the power.

Solution for (a)

Entering the known values of current and voltage for the hot headlight, we obtain

Equation:

$$P = IV = (2.50 \text{ A})(12.0 \text{ V}) = 30.0 \text{ W}.$$

The cold resistance was $0.350\ \Omega$, and so the power it uses when first switched on is

Equation:

$$P = \frac{V^2}{R} = \frac{(12.0\ \text{V})^2}{0.350\ \Omega} = 411\ \text{W}.$$

Discussion for (a)

The 30 W dissipated by the hot headlight is typical. But the 411 W when cold is surprisingly higher. The initial power quickly decreases as the bulb's temperature increases and its resistance increases.

Strategy and Solution for (b)

The current when the bulb is cold can be found several different ways. We rearrange one of the power equations, $P = I^2 R$, and enter known values, obtaining

Equation:

$$I = \sqrt{\frac{P}{R}} = \sqrt{\frac{411\ \text{W}}{0.350\ \Omega}} = 34.3\ \text{A}.$$

Discussion for (b)

The cold current is remarkably higher than the steady-state value of 2.50 A, but the current will quickly decline to that value as the bulb's temperature increases. Most fuses and circuit breakers (used to limit the current in a circuit) are designed to tolerate very high currents briefly as a device comes on. In some cases, such as with electric motors, the current remains high for several seconds, necessitating special "slow blow" fuses.

The Cost of Electricity

The more electric appliances you use and the longer they are left on, the higher your electric bill. This familiar fact is based on the relationship between energy and power. You pay for the energy used. Since $P = E/t$, we see that

Equation:

$$E = Pt$$

is the energy used by a device using power P for a time interval t . For example, the more lightbulbs burning, the greater P used; the longer they are on, the greater t is. The energy unit on electric bills is the kilowatt-hour ($\text{kW} \cdot \text{h}$), consistent with the relationship $E = Pt$. It is easy to estimate the cost of operating electric appliances if you have some idea of their power consumption rate in watts or kilowatts, the time they are on in hours, and the cost per kilowatt-hour for your electric utility. Kilowatt-hours, like all other specialized energy units such as food calories, can be converted to joules. You can prove to yourself that $1 \text{ kW} \cdot \text{h} = 3.6 \times 10^6 \text{ J}$.

The electrical energy (E) used can be reduced either by reducing the time of use or by reducing the power consumption of that appliance or fixture. This will not only reduce the cost, but it will also result in a reduced impact on the environment. Improvements to lighting are some of the fastest ways to reduce the electrical energy used in a home or business. About 20% of a home's use of energy goes to lighting, while the number for commercial establishments is closer to 40%. Fluorescent lights are about four times more efficient than incandescent lights—this is true for both the long tubes and the compact fluorescent lights (CFL). (See [\[link\]](#)(b).) Thus, a 60-W incandescent bulb can be replaced by a 15-W CFL, which has the same brightness and color. CFLs have a bent tube inside a globe or a spiral-shaped tube, all connected to a standard screw-in base that fits standard incandescent light sockets. (Original problems with color, flicker, shape, and high initial investment for CFLs have been addressed in recent years.) The heat transfer from these CFLs is less, and they last up to 10 times longer. The significance of an investment in such bulbs is addressed in the next example. New white LED lights (which are clusters of small LED bulbs) are even more efficient (twice that of CFLs) and last 5 times longer than CFLs. However, their cost is still high.

Note:

Making Connections: Energy, Power, and Time

The relationship $E = Pt$ is one that you will find useful in many different contexts. The energy your body uses in exercise is related to the power level and duration of your activity, for example. The amount of heating by a power source is related to the power level and time it is applied. Even the radiation dose of an X-ray image is related to the power and time of exposure.

Example:

Calculating the Cost Effectiveness of Compact Fluorescent Lights (CFL)

If the cost of electricity in your area is 12 cents per kWh, what is the total cost (capital plus operation) of using a 60-W incandescent bulb for 1000 hours (the lifetime of that bulb) if the bulb cost 25 cents? (b) If we replace this bulb with a compact fluorescent light that provides the same light output, but at one-quarter the wattage, and which costs \$1.50 but lasts 10 times longer (10,000 hours), what will that total cost be?

Strategy

To find the operating cost, we first find the energy used in kilowatt-hours and then multiply by the cost per kilowatt-hour.

Solution for (a)

The energy used in kilowatt-hours is found by entering the power and time into the expression for energy:

Equation:

$$E = Pt = (60 \text{ W})(1000 \text{ h}) = 60,000 \text{ W} \cdot \text{h}.$$

In kilowatt-hours, this is

Equation:

$$E = 60.0 \text{ kW} \cdot \text{h}.$$

Now the electricity cost is

Equation:

$$\text{cost} = (60.0 \text{ kW} \cdot \text{h})(\$0.12/\text{kW} \cdot \text{h}) = \$7.20.$$

The total cost will be \$7.20 for 1000 hours (about one-half year at 5 hours per day).

Solution for (b)

Since the CFL uses only 15 W and not 60 W, the electricity cost will be $\$7.20/4 = \1.80 . The CFL will last 10 times longer than the incandescent, so that the investment cost will be 1/10 of the bulb cost for that time period of use, or $0.1(\$1.50) = \0.15 . Therefore, the total cost will be \$1.95 for 1000 hours.

Discussion

Therefore, it is much cheaper to use the CFLs, even though the initial investment is higher. The increased cost of labor that a business must include for replacing the incandescent bulbs more often has not been figured in here.

Note:

Making Connections: Take-Home Experiment—Electrical Energy Use Inventory

1) Make a list of the power ratings on a range of appliances in your home or room. Explain why something like a toaster has a higher rating than a digital clock. Estimate the energy consumed by these appliances in an average day (by estimating their time of use). Some appliances might only state the operating current. If the household voltage is 120 V, then use $P = IV$. 2) Check out the total wattage used in the rest rooms of your school's floor or building. (You might need to assume the long fluorescent lights in use are rated at 32 W.) Suppose that the building was closed all weekend and that these lights were left on from 6 p.m. Friday until 8 a.m. Monday. What would this oversight cost? How about for an entire year of weekends?

Section Summary

- Electric power P is the rate (in watts) that energy is supplied by a source or dissipated by a device.

- Three expressions for electrical power are
Equation:

$$P = IV,$$

Equation:

$$P = \frac{V^2}{R},$$

and

Equation:

$$P = I^2 R.$$

- The energy used by a device with a power P over a time t is $E = Pt$.

Conceptual Questions

Exercise:

Problem:

Why do incandescent lightbulbs grow dim late in their lives, particularly just before their filaments break?

Exercise:

Problem:

The power dissipated in a resistor is given by $P = V^2/R$, which means power decreases if resistance increases. Yet this power is also given by $P = I^2 R$, which means power increases if resistance increases. Explain why there is no contradiction here.

Problem Exercises

Exercise:

Problem:

What is the power of a 1.00×10^2 MV lightning bolt having a current of 2.00×10^4 A?

Solution:

$$2.00 \times 10^{12} \text{ W}$$

Exercise:**Problem:**

What power is supplied to the starter motor of a large truck that draws 250 A of current from a 24.0-V battery hookup?

Exercise:**Problem:**

A charge of 4.00 C of charge passes through a pocket calculator's solar cells in 4.00 h. What is the power output, given the calculator's voltage output is 3.00 V? (See [\[link\]](#).)



The strip of solar cells just above the keys of this calculator convert

light to electricity
to supply its energy
needs. (credit:
Evan-Amos,
Wikimedia
Commons)

Exercise:

Problem:

How many watts does a flashlight that has 6.00×10^2 C pass through it in 0.500 h use if its voltage is 3.00 V?

Exercise:

Problem:

Find the power dissipated in each of these extension cords: (a) an extension cord having a $0.0600\ \Omega$ resistance and through which 5.00 A is flowing; (b) a cheaper cord utilizing thinner wire and with a resistance of $0.300\ \Omega$.

Solution:

(a) 1.50 W

(b) 7.50 W

Exercise:

Problem:

Verify that the units of a volt-ampere are watts, as implied by the equation $P = IV$.

Exercise:

Problem:

Show that the units $1 \text{ V}^2/\Omega = 1 \text{ W}$, as implied by the equation $P = V^2/R$.

Solution:

$$\frac{V^2}{\Omega} = \frac{V^2}{V/A} = AV = \left(\frac{C}{s}\right)\left(\frac{J}{C}\right) = \frac{J}{s} = 1 \text{ W}$$

Exercise:**Problem:**

Show that the units $1 \text{ A}^2 \cdot \Omega = 1 \text{ W}$, as implied by the equation $P = I^2 R$.

Exercise:**Problem:**

Verify the energy unit equivalence that $1 \text{ kW} \cdot \text{h} = 3.60 \times 10^6 \text{ J}$.

Solution:

$$1 \text{ kW} \cdot \text{h} = \left(\frac{1 \times 10^3 \text{ J}}{1 \text{ s}}\right)(1 \text{ h})\left(\frac{3600 \text{ s}}{1 \text{ h}}\right) = 3.60 \times 10^6 \text{ J}$$

Exercise:**Problem:**

Electrons in an X-ray tube are accelerated through $1.00 \times 10^2 \text{ kV}$ and directed toward a target to produce X-rays. Calculate the power of the electron beam in this tube if it has a current of 15.0 mA .

Exercise:

Problem:

An electric water heater consumes 5.00 kW for 2.00 h per day. What is the cost of running it for one year if electricity costs 12.0 cents/kW · h? See [\[link\]](#).



On-demand electric hot water heater. Heat is supplied to water only when needed.
(credit: aviddavid, Flickr)

Solution:

\$438/y

Exercise:**Problem:**

With a 1200-W toaster, how much electrical energy is needed to make a slice of toast (cooking time = 1 minute)? At 9.0 cents/kW · h, how much does this cost?

Exercise:

Problem:

What would be the maximum cost of a CFL such that the total cost (investment plus operating) would be the same for both CFL and incandescent 60-W bulbs? Assume the cost of the incandescent bulb is 25 cents and that electricity costs 10 cents/kWh. Calculate the cost for 1000 hours, as in the cost effectiveness of CFL example.

Solution:

\$6.25

Exercise:**Problem:**

Some makes of older cars have 6.00-V electrical systems. (a) What is the hot resistance of a 30.0-W headlight in such a car? (b) What current flows through it?

Exercise:**Problem:**

Alkaline batteries have the advantage of putting out constant voltage until very nearly the end of their life. How long will an alkaline battery rated at 1.00 A · h and 1.58 V keep a 1.00-W flashlight bulb burning?

Solution:

1.58 h

Exercise:**Problem:**

A cauterizer, used to stop bleeding in surgery, puts out 2.00 mA at 15.0 kV. (a) What is its power output? (b) What is the resistance of the path?

Exercise:

Problem:

The average television is said to be on 6 hours per day. Estimate the yearly cost of electricity to operate 100 million TVs, assuming their power consumption averages 150 W and the cost of electricity averages 12.0 cents/kW · h.

Solution:

\$3.94 billion/year

Exercise:**Problem:**

An old lightbulb draws only 50.0 W, rather than its original 60.0 W, due to evaporative thinning of its filament. By what factor is its diameter reduced, assuming uniform thinning along its length? Neglect any effects caused by temperature differences.

Exercise:**Problem:**

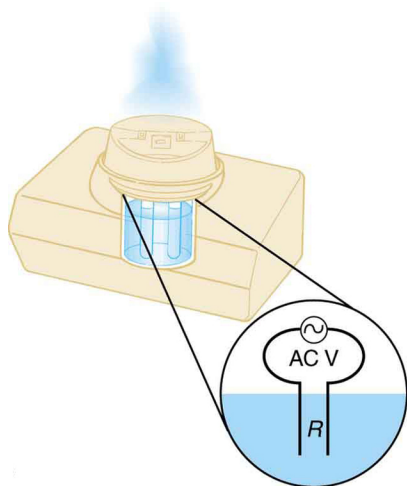
00-gauge copper wire has a diameter of 9.266 mm. Calculate the power loss in a kilometer of such wire when it carries 1.00×10^2 A.

Solution:

25.5 W

Exercise:**Problem: Integrated Concepts**

Cold vaporizers pass a current through water, evaporating it with only a small increase in temperature. One such home device is rated at 3.50 A and utilizes 120 V AC with 95.0% efficiency. (a) What is the vaporization rate in grams per minute? (b) How much water must you put into the vaporizer for 8.00 h of overnight operation? (See [\[link\]](#).)



This cold vaporizer
passes current
directly through
water, vaporizing it
directly with
relatively little
temperature
increase.

Exercise:

Problem: Integrated Concepts

(a) What energy is dissipated by a lightning bolt having a 20,000-A current, a voltage of 1.00×10^2 MV, and a length of 1.00 ms? (b) What mass of tree sap could be raised from 18.0°C to its boiling point and then evaporated by this energy, assuming sap has the same thermal characteristics as water?

Solution:

(a) 2.00×10^9 J

(b) 769 kg

Exercise:**Problem: Integrated Concepts**

What current must be produced by a 12.0-V battery-operated bottle warmer in order to heat 75.0 g of glass, 250 g of baby formula, and 3.00×10^2 g of aluminum from 20.0°C to 90.0°C in 5.00 min?

Exercise:**Problem: Integrated Concepts**

How much time is needed for a surgical cauterizer to raise the temperature of 1.00 g of tissue from 37.0°C to 100°C and then boil away 0.500 g of water, if it puts out 2.00 mA at 15.0 kV? Ignore heat transfer to the surroundings.

Solution:

45.0 s

Exercise:**Problem: Integrated Concepts**

Hydroelectric generators (see [\[link\]](#)) at Hoover Dam produce a maximum current of 8.00×10^3 A at 250 kV. (a) What is the power output? (b) The water that powers the generators enters and leaves the system at low speed (thus its kinetic energy does not change) but loses 160 m in altitude. How many cubic meters per second are needed, assuming 85.0% efficiency?



Hydroelectric generators
at the Hoover dam.
(credit: Jon Sullivan)

Exercise:

Problem: Integrated Concepts

(a) Assuming 95.0% efficiency for the conversion of electrical power by the motor, what current must the 12.0-V batteries of a 750-kg electric car be able to supply: (a) To accelerate from rest to 25.0 m/s in 1.00 min? (b) To climb a 2.00×10^2 -m-high hill in 2.00 min at a constant 25.0-m/s speed while exerting 5.00×10^2 N of force to overcome air resistance and friction? (c) To travel at a constant 25.0-m/s speed, exerting a 5.00×10^2 N force to overcome air resistance and friction? See [\[link\]](#).



This REVAi, an electric

car, gets recharged on a street in London. (credit: Frank Hebbert)

Solution:

(a) 343 A

(b) 2.17×10^3 A

(c) 1.10×10^3 A

Exercise:

Problem: Integrated Concepts

A light-rail commuter train draws 630 A of 650-V DC electricity when accelerating. (a) What is its power consumption rate in kilowatts? (b) How long does it take to reach 20.0 m/s starting from rest if its loaded mass is 5.30×10^4 kg, assuming 95.0% efficiency and constant power? (c) Find its average acceleration. (d) Discuss how the acceleration you found for the light-rail train compares to what might be typical for an automobile.

Exercise:

Problem: Integrated Concepts

(a) An aluminum power transmission line has a resistance of $0.0580 \Omega/\text{km}$. What is its mass per kilometer? (b) What is the mass per kilometer of a copper line having the same resistance? A lower resistance would shorten the heating time. Discuss the practical limits to speeding the heating by lowering the resistance.

Solution:

(a) $1.23 \times 10^3 \text{ kg}$

(b) $2.64 \times 10^3 \text{ kg}$

Exercise:

Problem: Integrated Concepts

(a) An immersion heater utilizing 120 V can raise the temperature of a 1.00×10^2 -g aluminum cup containing 350 g of water from 20.0°C to 95.0°C in 2.00 min. Find its resistance, assuming it is constant during the process. (b) A lower resistance would shorten the heating time. Discuss the practical limits to speeding the heating by lowering the resistance.

Exercise:

Problem: Integrated Concepts

(a) What is the cost of heating a hot tub containing 1500 kg of water from 10.0°C to 40.0°C, assuming 75.0% efficiency to account for heat transfer to the surroundings? The cost of electricity is 9 cents/kW · h. (b) What current was used by the 220-V AC electric heater, if this took 4.00 h?

Exercise:

Problem: Unreasonable Results

(a) What current is needed to transmit 1.00×10^2 MW of power at 480 V? (b) What power is dissipated by the transmission lines if they have a $1.00 - \Omega$ resistance? (c) What is unreasonable about this result? (d) Which assumptions are unreasonable, or which premises are inconsistent?

Solution:

(a) $2.08 \times 10^5 \text{ A}$

(b) 4.33×10^4 MW

(c) The transmission lines dissipate more power than they are supposed to transmit.

(d) A voltage of 480 V is unreasonably low for a transmission voltage. Long-distance transmission lines are kept at much higher voltages (often hundreds of kilovolts) to reduce power losses.

Exercise:

Problem: Unreasonable Results

(a) What current is needed to transmit 1.00×10^2 MW of power at 10.0 kV? (b) Find the resistance of 1.00 km of wire that would cause a 0.0100% power loss. (c) What is the diameter of a 1.00-km-long copper wire having this resistance? (d) What is unreasonable about these results? (e) Which assumptions are unreasonable, or which premises are inconsistent?

Exercise:

Problem: Construct Your Own Problem

Consider an electric immersion heater used to heat a cup of water to make tea. Construct a problem in which you calculate the needed resistance of the heater so that it increases the temperature of the water and cup in a reasonable amount of time. Also calculate the cost of the electrical energy used in your process. Among the things to be considered are the voltage used, the masses and heat capacities involved, heat losses, and the time over which the heating takes place. Your instructor may wish for you to consider a thermal safety switch (perhaps bimetallic) that will halt the process before damaging temperatures are reached in the immersion unit.

Glossary

electric power

the rate at which electrical energy is supplied by a source or dissipated by a device; it is the product of current times voltage

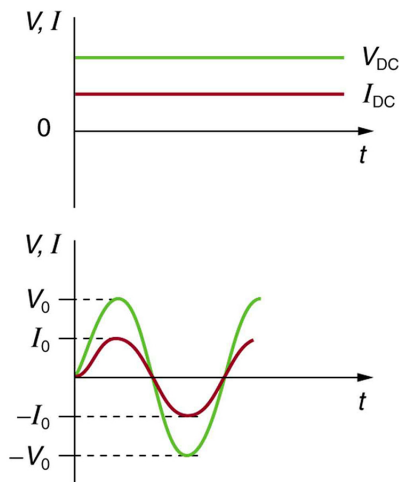
Alternating Current versus Direct Current

- Explain the differences and similarities between AC and DC current.
- Calculate rms voltage, current, and average power.
- Explain why AC current is used for power transmission.

Alternating Current

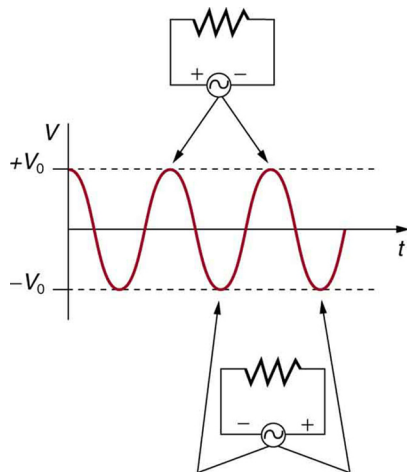
Most of the examples dealt with so far, and particularly those utilizing batteries, have constant voltage sources. Once the current is established, it is thus also a constant. **Direct current** (DC) is the flow of electric charge in only one direction. It is the steady state of a constant-voltage circuit. Most well-known applications, however, use a time-varying voltage source.

Alternating current (AC) is the flow of electric charge that periodically reverses direction. If the source varies periodically, particularly sinusoidally, the circuit is known as an alternating current circuit. Examples include the commercial and residential power that serves so many of our needs. [\[link\]](#) shows graphs of voltage and current versus time for typical DC and AC power. The AC voltages and frequencies commonly used in homes and businesses vary around the world.



(a) DC voltage and current are constant in time, once the

current is established. (b) A graph of voltage and current versus time for 60-Hz AC power. The voltage and current are sinusoidal and are in phase for a simple resistance circuit. The frequencies and peak voltages of AC sources differ greatly.



The potential difference V between the terminals of an AC voltage source fluctuates as

shown. The
mathematical
expression for V is
given by
 $V = V_0 \sin 2\pi ft$.

[\[link\]](#) shows a schematic of a simple circuit with an AC voltage source. The voltage between the terminals fluctuates as shown, with the **AC voltage** given by

Equation:

$$V = V_0 \sin 2\pi ft,$$

where V is the voltage at time t , V_0 is the peak voltage, and f is the frequency in hertz. For this simple resistance circuit, $I = V/R$, and so the **AC current** is

Equation:

$$I = I_0 \sin 2\pi ft,$$

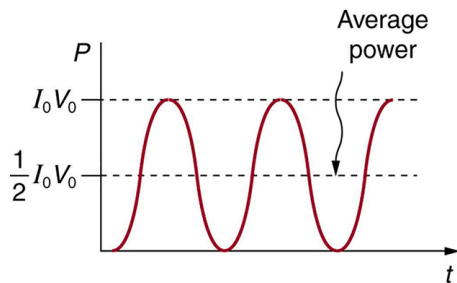
where I is the current at time t , and $I_0 = V_0/R$ is the peak current. For this example, the voltage and current are said to be in phase, as seen in [\[link\]](#)(b).

Current in the resistor alternates back and forth just like the driving voltage, since $I = V/R$. If the resistor is a fluorescent light bulb, for example, it brightens and dims 120 times per second as the current repeatedly goes through zero. A 120-Hz flicker is too rapid for your eyes to detect, but if you wave your hand back and forth between your face and a fluorescent light, you will see a stroboscopic effect evidencing AC. The fact that the light output fluctuates means that the power is fluctuating. The power supplied is $P = IV$. Using the expressions for I and V above, we see that the time dependence of power is $P = I_0 V_0 \sin^2 2\pi ft$, as shown in [\[link\]](#).

Note:**Making Connections: Take-Home Experiment—AC/DC Lights**

Wave your hand back and forth between your face and a fluorescent light bulb. Do you observe the same thing with the headlights on your car?

Explain what you observe. *Warning: Do not look directly at very bright light.*



AC power as a function of time. Since the voltage and current are in phase here, their product is non-negative and fluctuates between zero and $I_0 V_0$. Average power is $(1/2) I_0 V_0$.

We are most often concerned with average power rather than its fluctuations—that 60-W light bulb in your desk lamp has an average power consumption of 60 W, for example. As illustrated in [\[link\]](#), the average power P_{ave} is

Equation:

$$P_{\text{ave}} = \frac{1}{2} I_0 V_0.$$

This is evident from the graph, since the areas above and below the $(1/2)I_0V_0$ line are equal, but it can also be proven using trigonometric identities. Similarly, we define an average or **rms current** I_{rms} and average or **rms voltage** V_{rms} to be, respectively,

Equation:

$$I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$$

and

Equation:

$$V_{\text{rms}} = \frac{V_0}{\sqrt{2}}.$$

where rms stands for root mean square, a particular kind of average. In general, to obtain a root mean square, the particular quantity is squared, its mean (or average) is found, and the square root is taken. This is useful for AC, since the average value is zero. Now,

Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}},$$

which gives

Equation:

$$P_{\text{ave}} = \frac{I_0}{\sqrt{2}} \cdot \frac{V_0}{\sqrt{2}} = \frac{1}{2} I_0 V_0,$$

as stated above. It is standard practice to quote I_{rms} , V_{rms} , and P_{ave} rather than the peak values. For example, most household electricity is 120 V AC, which means that V_{rms} is 120 V. The common 10-A circuit breaker will interrupt a sustained I_{rms} greater than 10 A. Your 1.0-kW microwave oven

consumes $P_{\text{ave}} = 1.0 \text{ kW}$, and so on. You can think of these rms and average values as the equivalent DC values for a simple resistive circuit.

To summarize, when dealing with AC, Ohm's law and the equations for power are completely analogous to those for DC, but rms and average values are used for AC. Thus, for AC, Ohm's law is written

Equation:

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{R}.$$

The various expressions for AC power P_{ave} are

Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}},$$

Equation:

$$P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R},$$

and

Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2 R.$$

Example:

Peak Voltage and Power for AC

(a) What is the value of the peak voltage for 120-V AC power? (b) What is the peak power consumption rate of a 60.0-W AC light bulb?

Strategy

We are told that V_{rms} is 120 V and P_{ave} is 60.0 W. We can use $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$ to find the peak voltage, and we can manipulate the definition of power to

find the peak power from the given average power.

Solution for (a)

Solving the equation $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$ for the peak voltage V_0 and substituting the known value for V_{rms} gives

Equation:

$$V_0 = \sqrt{2}V_{\text{rms}} = 1.414(120 \text{ V}) = 170 \text{ V}.$$

Discussion for (a)

This means that the AC voltage swings from 170 V to -170 V and back 60 times every second. An equivalent DC voltage is a constant 120 V.

Solution for (b)

Peak power is peak current times peak voltage. Thus,

Equation:

$$P_0 = I_0 V_0 = 2 \left(\frac{1}{2} I_0 V_0 \right) = 2P_{\text{ave}}.$$

We know the average power is 60.0 W, and so

Equation:

$$P_0 = 2(60.0 \text{ W}) = 120 \text{ W}.$$

Discussion

So the power swings from zero to 120 W one hundred twenty times per second (twice each cycle), and the power averages 60 W.

Why Use AC for Power Distribution?

Most large power-distribution systems are AC. Moreover, the power is transmitted at much higher voltages than the 120-V AC (240 V in most parts of the world) we use in homes and on the job. Economies of scale make it cheaper to build a few very large electric power-generation plants than to build numerous small ones. This necessitates sending power long distances, and it is obviously important that energy losses en route be

minimized. High voltages can be transmitted with much smaller power losses than low voltages, as we shall see. (See [\[link\]](#).) For safety reasons, the voltage at the user is reduced to familiar values. The crucial factor is that it is much easier to increase and decrease AC voltages than DC, so AC is used in most large power distribution systems.



Power is distributed over large distances at high voltage to reduce power loss in the transmission lines. The voltages generated at the power plant are stepped up by passive devices called transformers (see [Transformers](#)) to 330,000 volts (or more in some places worldwide). At the point of use, the transformers reduce the voltage transmitted for safe residential and commercial use.

(Credit: GeorgHH, Wikimedia Commons)

Example:**Power Losses Are Less for High-Voltage Transmission**

(a) What current is needed to transmit 100 MW of power at 200 kV? (b) What is the power dissipated by the transmission lines if they have a resistance of $1.00\ \Omega$? (c) What percentage of the power is lost in the transmission lines?

Strategy

We are given $P_{\text{ave}} = 100\text{ MW}$, $V_{\text{rms}} = 200\text{ kV}$, and the resistance of the lines is $R = 1.00\ \Omega$. Using these givens, we can find the current flowing (from $P = IV$) and then the power dissipated in the lines ($P = I^2 R$), and we take the ratio to the total power transmitted.

Solution

To find the current, we rearrange the relationship $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$ and substitute known values. This gives

Equation:

$$I_{\text{rms}} = \frac{P_{\text{ave}}}{V_{\text{rms}}} = \frac{100 \times 10^6\text{ W}}{200 \times 10^3\text{ V}} = 500\text{ A}.$$

Solution

Knowing the current and given the resistance of the lines, the power dissipated in them is found from $P_{\text{ave}} = I_{\text{rms}}^2 R$. Substituting the known values gives

Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2 R = (500\text{ A})^2 (1.00\ \Omega) = 250\text{ kW}.$$

Solution

The percent loss is the ratio of this lost power to the total or input power, multiplied by 100:

Equation:

$$\% \text{ loss} = \frac{250 \text{ kW}}{100 \text{ MW}} \times 100 = 0.250 \%$$

Discussion

One-fourth of a percent is an acceptable loss. Note that if 100 MW of power had been transmitted at 25 kV, then a current of 4000 A would have been needed. This would result in a power loss in the lines of 16.0 MW, or 16.0% rather than 0.250%. The lower the voltage, the more current is needed, and the greater the power loss in the fixed-resistance transmission lines. Of course, lower-resistance lines can be built, but this requires larger and more expensive wires. If superconducting lines could be economically produced, there would be no loss in the transmission lines at all. But, as we shall see in a later chapter, there is a limit to current in superconductors, too. In short, high voltages are more economical for transmitting power, and AC voltage is much easier to raise and lower, so that AC is used in most large-scale power distribution systems.

It is widely recognized that high voltages pose greater hazards than low voltages. But, in fact, some high voltages, such as those associated with common static electricity, can be harmless. So it is not voltage alone that determines a hazard. It is not so widely recognized that AC shocks are often more harmful than similar DC shocks. Thomas Edison thought that AC shocks were more harmful and set up a DC power-distribution system in New York City in the late 1800s. There were bitter fights, in particular between Edison and George Westinghouse and Nikola Tesla, who were advocating the use of AC in early power-distribution systems. AC has prevailed largely due to transformers and lower power losses with high-voltage transmission.

Note:**PhET Explorations: Generator**

Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

Generato

r

Section Summary

- Direct current (DC) is the flow of electric current in only one direction. It refers to systems where the source voltage is constant.
- The voltage source of an alternating current (AC) system puts out $V = V_0 \sin 2\pi ft$, where V is the voltage at time t , V_0 is the peak voltage, and f is the frequency in hertz.
- In a simple circuit, $I = V/R$ and AC current is $I = I_0 \sin 2\pi ft$, where I is the current at time t , and $I_0 = V_0/R$ is the peak current.
- The average AC power is $P_{\text{ave}} = \frac{1}{2} I_0 V_0$.
- Average (rms) current I_{rms} and average (rms) voltage V_{rms} are $I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$ and $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$, where rms stands for root mean square.
- Thus, $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$.
- Ohm's law for AC is $I_{\text{rms}} = \frac{V_{\text{rms}}}{R}$.
- Expressions for the average power of an AC circuit are $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$, $P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R}$, and $P_{\text{ave}} = I_{\text{rms}}^2 R$, analogous to the expressions for DC circuits.

Conceptual Questions

Exercise:

Problem:

Give an example of a use of AC power other than in the household. Similarly, give an example of a use of DC power other than that supplied by batteries.

Exercise:**Problem:**

Why do voltage, current, and power go through zero 120 times per second for 60-Hz AC electricity?

Exercise:**Problem:**

You are riding in a train, gazing into the distance through its window. As close objects streak by, you notice that the nearby fluorescent lights make *dashed* streaks. Explain.

Problem Exercises**Exercise:****Problem:**

(a) What is the hot resistance of a 25-W light bulb that runs on 120-V AC? (b) If the bulb's operating temperature is 2700°C , what is its resistance at 2600°C ?

Exercise:**Problem:**

Certain heavy industrial equipment uses AC power that has a peak voltage of 679 V. What is the rms voltage?

Solution:

480 V

Exercise:**Problem:**

A certain circuit breaker trips when the rms current is 15.0 A. What is the corresponding peak current?

Exercise:**Problem:**

Military aircraft use 400-Hz AC power, because it is possible to design lighter-weight equipment at this higher frequency. What is the time for one complete cycle of this power?

Solution:

2.50 ms

Exercise:**Problem:**

A North American tourist takes his 25.0-W, 120-V AC razor to Europe, finds a special adapter, and plugs it into 240 V AC. Assuming constant resistance, what power does the razor consume as it is ruined?

Exercise:**Problem:**

In this problem, you will verify statements made at the end of the power losses for [link](#). (a) What current is needed to transmit 100 MW of power at a voltage of 25.0 kV? (b) Find the power loss in a 1.00 - Ω transmission line. (c) What percent loss does this represent?

Solution:

(a) 4.00 kA

(b) 16.0 MW

(c) 16.0%

Exercise:

Problem:

A small office-building air conditioner operates on 408-V AC and consumes 50.0 kW. (a) What is its effective resistance? (b) What is the cost of running the air conditioner during a hot summer month when it is on 8.00 h per day for 30 days and electricity costs 9.00 cents/kW · h?

Exercise:**Problem:**

What is the peak power consumption of a 120-V AC microwave oven that draws 10.0 A?

Solution:

2.40 kW

Exercise:**Problem:**

What is the peak current through a 500-W room heater that operates on 120-V AC power?

Exercise:**Problem:**

Two different electrical devices have the same power consumption, but one is meant to be operated on 120-V AC and the other on 240-V AC. (a) What is the ratio of their resistances? (b) What is the ratio of their currents? (c) Assuming its resistance is unaffected, by what factor will the power increase if a 120-V AC device is connected to 240-V AC?

Solution:

(a) 4.0

(b) 0.50

(c) 4.0

Exercise:

Problem:

Nichrome wire is used in some radiative heaters. (a) Find the resistance needed if the average power output is to be 1.00 kW utilizing 120-V AC. (b) What length of Nichrome wire, having a cross-sectional area of 5.00mm^2 , is needed if the operating temperature is 500°C ? (c) What power will it draw when first switched on?

Exercise:

Problem:

Find the time after $t = 0$ when the instantaneous voltage of 60-Hz AC first reaches the following values: (a) $V_0/2$ (b) V_0 (c) 0.

Solution:

(a) 1.39 ms

(b) 4.17 ms

(c) 8.33 ms

Exercise:

Problem:

(a) At what two times in the first period following $t = 0$ does the instantaneous voltage in 60-Hz AC equal V_{rms} ? (b) $-V_{\text{rms}}$?

Glossary

direct current

(DC) the flow of electric charge in only one direction

alternating current

(AC) the flow of electric charge that periodically reverses direction

AC voltage

voltage that fluctuates sinusoidally with time, expressed as $V = V_0 \sin 2\pi ft$, where V is the voltage at time t , V_0 is the peak voltage, and f is the frequency in hertz

AC current

current that fluctuates sinusoidally with time, expressed as $I = I_0 \sin 2\pi ft$, where I is the current at time t , I_0 is the peak current, and f is the frequency in hertz

rms current

the root mean square of the current, $I_{\text{rms}} = I_0/\sqrt{2}$, where I_0 is the peak current, in an AC system

rms voltage

the root mean square of the voltage, $V_{\text{rms}} = V_0/\sqrt{2}$, where V_0 is the peak voltage, in an AC system

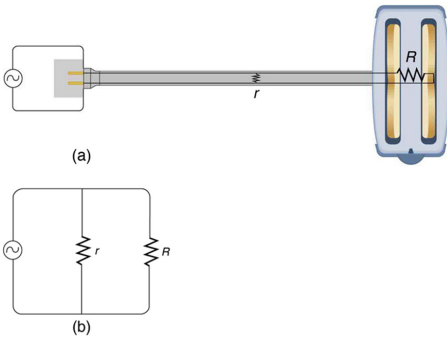
Electric Hazards and the Human Body

- Define thermal hazard, shock hazard, and short circuit.
- Explain what effects various levels of current have on the human body.

There are two known hazards of electricity—thermal and shock. A **thermal hazard** is one where excessive electric power causes undesired thermal effects, such as starting a fire in the wall of a house. A **shock hazard** occurs when electric current passes through a person. Shocks range in severity from painful, but otherwise harmless, to heart-stopping lethality. This section considers these hazards and the various factors affecting them in a quantitative manner. [Electrical Safety: Systems and Devices](#) will consider systems and devices for preventing electrical hazards.

Thermal Hazards

Electric power causes undesired heating effects whenever electric energy is converted to thermal energy at a rate faster than it can be safely dissipated. A classic example of this is the **short circuit**, a low-resistance path between terminals of a voltage source. An example of a short circuit is shown in [\[link\]](#). Insulation on wires leading to an appliance has worn through, allowing the two wires to come into contact. Such an undesired contact with a high voltage is called a *short*. Since the resistance of the short, r , is very small, the power dissipated in the short, $P = V^2/r$, is very large. For example, if V is 120 V and r is 0.100 Ω , then the power is 144 kW, *much* greater than that used by a typical household appliance. Thermal energy delivered at this rate will very quickly raise the temperature of surrounding materials, melting or perhaps igniting them.

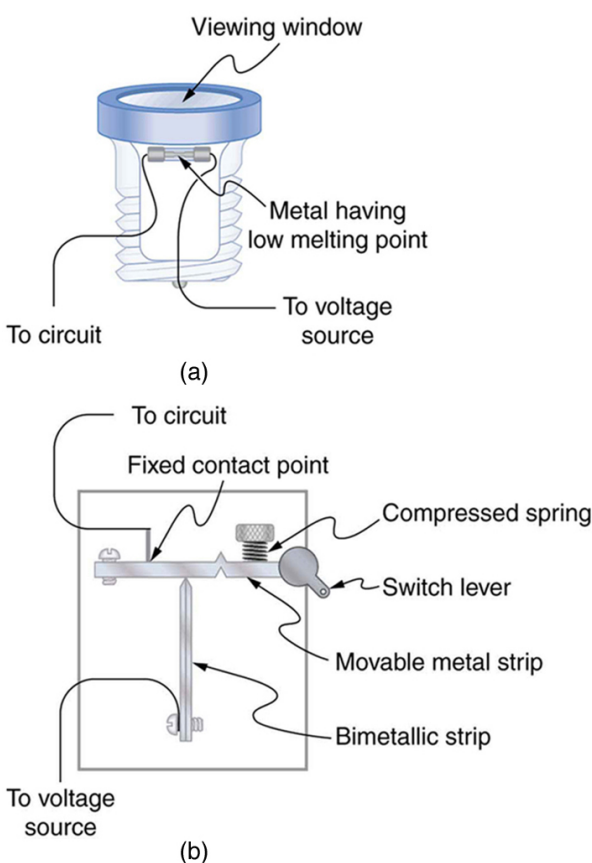


A short circuit is an undesired low-resistance path across a voltage source. (a) Worn insulation on the wires of a toaster allow them to come into contact with a low resistance r . Since $P = V^2/r$, thermal power is created so rapidly that the cord melts or burns. (b) A schematic of the short circuit.

One particularly insidious aspect of a short circuit is that its resistance may actually be decreased due to the increase in temperature. This can happen if the short creates ionization. These charged atoms and molecules are free to move and, thus, lower the resistance r . Since $P = V^2/r$, the power dissipated in the short rises, possibly causing more ionization, more power, and so on. High voltages, such as the 480-V AC used in some industrial applications, lend themselves to this hazard, because higher voltages create higher initial power production in a short.

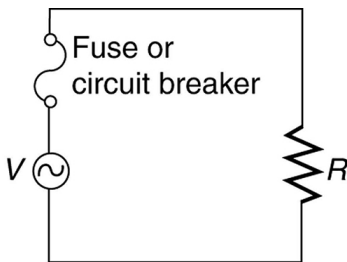
Another serious, but less dramatic, thermal hazard occurs when wires supplying power to a user are overloaded with too great a current. As

discussed in the previous section, the power dissipated in the supply wires is $P = I^2 R_w$, where R_w is the resistance of the wires and I the current flowing through them. If either I or R_w is too large, the wires overheat. For example, a worn appliance cord (with some of its braided wires broken) may have $R_w = 2.00 \, \Omega$ rather than the $0.100 \, \Omega$ it should be. If $10.0 \, \text{A}$ of current passes through the cord, then $P = I^2 R_w = 200 \, \text{W}$ is dissipated in the cord—much more than is safe. Similarly, if a wire with a $0.100 \, \Omega$ resistance is meant to carry a few amps, but is instead carrying $100 \, \text{A}$, it will severely overheat. The power dissipated in the wire will in that case be $P = 1000 \, \text{W}$. Fuses and circuit breakers are used to limit excessive currents. (See [\[link\]](#) and [\[link\]](#).) Each device opens the circuit automatically when a sustained current exceeds safe limits.



(a) A fuse has a metal strip with a low melting point that, when overheated by an excessive

current, permanently breaks the connection of a circuit to a voltage source. (b) A circuit breaker is an automatic but restorable electric switch. The one shown here has a bimetallic strip that bends to the right and into the notch if overheated. The spring then forces the metal strip downward, breaking the electrical connection at the points.



Schematic of a circuit with a fuse or circuit breaker in it.

Fuses and circuit breakers act like automatic switches that open when sustained current exceeds desired limits.

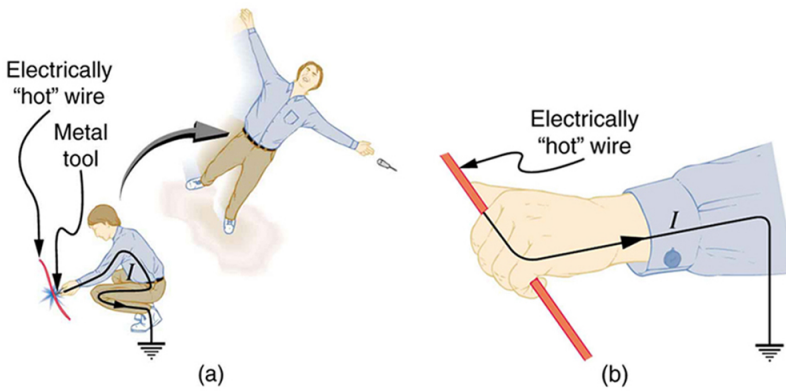
Fuses and circuit breakers for typical household voltages and currents are relatively simple to produce, but those for large voltages and currents experience special problems. For example, when a circuit breaker tries to interrupt the flow of high-voltage electricity, a spark can jump across its points that ionizes the air in the gap and allows the current to continue flowing. Large circuit breakers found in power-distribution systems employ insulating gas and even use jets of gas to blow out such sparks. Here AC is safer than DC, since AC current goes through zero 120 times per second, giving a quick opportunity to extinguish these arcs.

Shock Hazards

Electrical currents through people produce tremendously varied effects. An electrical current can be used to block back pain. The possibility of using electrical current to stimulate muscle action in paralyzed limbs, perhaps allowing paraplegics to walk, is under study. TV dramatizations in which electrical shocks are used to bring a heart attack victim out of ventricular fibrillation (a massively irregular, often fatal, beating of the heart) are more than common. Yet most electrical shock fatalities occur because a current put the heart into fibrillation. A pacemaker uses electrical shocks to stimulate the heart to beat properly. Some fatal shocks do not produce burns, but warts can be safely burned off with electric current (though freezing using liquid nitrogen is now more common). Of course, there are consistent explanations for these disparate effects. The major factors upon which the effects of electrical shock depend are

1. The amount of current I
2. The path taken by the current
3. The duration of the shock
4. The frequency f of the current ($f = 0$ for DC)

[\[link\]](#) gives the effects of electrical shocks as a function of current for a typical accidental shock. The effects are for a shock that passes through the trunk of the body, has a duration of 1 s, and is caused by 60-Hz power.



An electric current can cause muscular contractions with varying effects. (a) The victim is “thrown” backward by involuntary muscle contractions that extend the legs and torso. (b) The victim can’t let go of the wire that is stimulating all the muscles in the hand. Those that close the fingers are stronger than those that open them.

Current (mA)	Effect
1	Threshold of sensation
5	Maximum harmless current
10–20	Onset of sustained muscular contraction; cannot let go for duration of shock; contraction of chest muscles may stop breathing during shock

Current (mA)	Effect
50	Onset of pain
100–300+	Ventricular fibrillation possible; often fatal
300	Onset of burns depending on concentration of current
6000 (6 A)	Onset of sustained ventricular contraction and respiratory paralysis; both cease when shock ends; heartbeat may return to normal; used to defibrillate the heart

Effects of Electrical Shock as a Function of Current^[footnote]

For an average male shocked through trunk of body for 1 s by 60-Hz AC. Values for females are 60–80% of those listed.

Our bodies are relatively good conductors due to the water in our bodies. Given that larger currents will flow through sections with lower resistance (to be further discussed in the next chapter), electric currents preferentially flow through paths in the human body that have a minimum resistance in a direct path to earth. The earth is a natural electron sink. Wearing insulating shoes, a requirement in many professions, prohibits a pathway for electrons by providing a large resistance in that path. Whenever working with high-power tools (drills), or in risky situations, ensure that you do not provide a pathway for current flow (especially through the heart).

Very small currents pass harmlessly and unfelt through the body. This happens to you regularly without your knowledge. The threshold of sensation is only 1 mA and, although unpleasant, shocks are apparently harmless for currents less than 5 mA. A great number of safety rules take the 5-mA value for the maximum allowed shock. At 10 to 20 mA and above, the current can stimulate sustained muscular contractions much as regular nerve impulses do. People sometimes say they were knocked across the room by a shock, but what really happened was that certain muscles

contracted, propelling them in a manner not of their own choosing. (See [\[link\]](#)(a).) More frightening, and potentially more dangerous, is the “can’t let go” effect illustrated in [\[link\]](#)(b). The muscles that close the fingers are stronger than those that open them, so the hand closes involuntarily on the wire shocking it. This can prolong the shock indefinitely. It can also be a danger to a person trying to rescue the victim, because the rescuer’s hand may close about the victim’s wrist. Usually the best way to help the victim is to give the fist a hard knock/blow/jar with an insulator or to throw an insulator at the fist. Modern electric fences, used in animal enclosures, are now pulsed on and off to allow people who touch them to get free, rendering them less lethal than in the past.

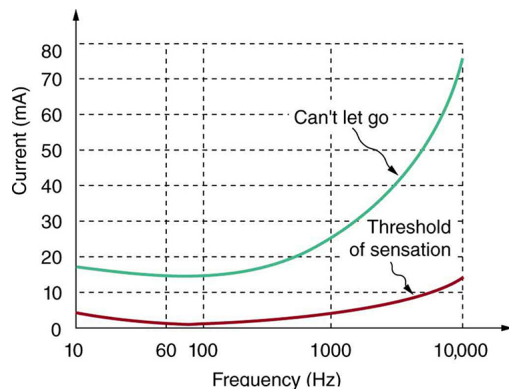
Greater currents may affect the heart. Its electrical patterns can be disrupted, so that it beats irregularly and ineffectively in a condition called “ventricular fibrillation.” This condition often lingers after the shock and is fatal due to a lack of blood circulation. The threshold for ventricular fibrillation is between 100 and 300 mA. At about 300 mA and above, the shock can cause burns, depending on the concentration of current—the more concentrated, the greater the likelihood of burns.

Very large currents cause the heart and diaphragm to contract for the duration of the shock. Both the heart and breathing stop. Interestingly, both often return to normal following the shock. The electrical patterns on the heart are completely erased in a manner that the heart can start afresh with normal beating, as opposed to the permanent disruption caused by smaller currents that can put the heart into ventricular fibrillation. The latter is something like scribbling on a blackboard, whereas the former completely erases it. TV dramatizations of electric shock used to bring a heart attack victim out of ventricular fibrillation also show large paddles. These are used to spread out current passed through the victim to reduce the likelihood of burns.

Current is the major factor determining shock severity (given that other conditions such as path, duration, and frequency are fixed, such as in the table and preceding discussion). A larger voltage is more hazardous, but since $I = V/R$, the severity of the shock depends on the combination of voltage and resistance. For example, a person with dry skin has a resistance

of about $200\text{ k}\Omega$. If he comes into contact with 120-V AC, a current $I = (120\text{ V})/(200\text{ k}\Omega) = 0.6\text{ mA}$ passes harmlessly through him. The same person soaking wet may have a resistance of $10.0\text{ k}\Omega$ and the same 120 V will produce a current of 12 mA—above the “can’t let go” threshold and potentially dangerous.

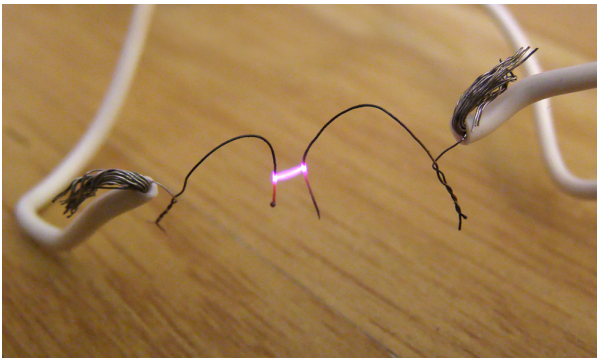
Most of the body’s resistance is in its dry skin. When wet, salts go into ion form, lowering the resistance significantly. The interior of the body has a much lower resistance than dry skin because of all the ionic solutions and fluids it contains. If skin resistance is bypassed, such as by an intravenous infusion, a catheter, or exposed pacemaker leads, a person is rendered **microshock sensitive**. In this condition, currents about 1/1000 those listed in [\[link\]](#) produce similar effects. During open-heart surgery, currents as small as $20\text{ }\mu\text{A}$ can be used to still the heart. Stringent electrical safety requirements in hospitals, particularly in surgery and intensive care, are related to the doubly disadvantaged microshock-sensitive patient. The break in the skin has reduced his resistance, and so the same voltage causes a greater current, and a much smaller current has a greater effect.



Graph of average values
for the threshold of
sensation and the “can’t
let go” current as a
function of frequency.
The lower the value, the

more sensitive the body is
at that frequency.

Factors other than current that affect the severity of a shock are its path, duration, and AC frequency. Path has obvious consequences. For example, the heart is unaffected by an electric shock through the brain, such as may be used to treat manic depression. And it is a general truth that the longer the duration of a shock, the greater its effects. [\[link\]](#) presents a graph that illustrates the effects of frequency on a shock. The curves show the minimum current for two different effects, as a function of frequency. The lower the current needed, the more sensitive the body is at that frequency. Ironically, the body is most sensitive to frequencies near the 50- or 60-Hz frequencies in common use. The body is slightly less sensitive for DC ($f = 0$), mildly confirming Edison's claims that AC presents a greater hazard. At higher and higher frequencies, the body becomes progressively less sensitive to any effects that involve nerves. This is related to the maximum rates at which nerves can fire or be stimulated. At very high frequencies, electrical current travels only on the surface of a person. Thus a wart can be burned off with very high frequency current without causing the heart to stop. (Do not try this at home with 60-Hz AC!) Some of the spectacular demonstrations of electricity, in which high-voltage arcs are passed through the air and over people's bodies, employ high frequencies and low currents. (See [\[link\]](#).) Electrical safety devices and techniques are discussed in detail in [Electrical Safety: Systems and Devices](#).



Is this electric arc dangerous?

The answer depends on the AC frequency and the power involved. (credit: Khimich Alex, Wikimedia Commons)

Section Summary

- The two types of electric hazards are thermal (excessive power) and shock (current through a person).
- Shock severity is determined by current, path, duration, and AC frequency.
- [\[link\]](#) lists shock hazards as a function of current.
- [\[link\]](#) graphs the threshold current for two hazards as a function of frequency.

Conceptual Questions

Exercise:

Problem:

Using an ohmmeter, a student measures the resistance between various points on his body. He finds that the resistance between two points on the same finger is about the same as the resistance between two points on opposite hands—both are several hundred thousand ohms. Furthermore, the resistance decreases when more skin is brought into contact with the probes of the ohmmeter. Finally, there is a dramatic drop in resistance (to a few thousand ohms) when the skin is wet. Explain these observations and their implications regarding skin and internal resistance of the human body.

Exercise:

Problem: What are the two major hazards of electricity?

Exercise:

Problem: Why isn't a short circuit a shock hazard?

Exercise:

Problem:

What determines the severity of a shock? Can you say that a certain voltage is hazardous without further information?

Exercise:

Problem:

An electrified needle is used to burn off warts, with the circuit being completed by having the patient sit on a large butt plate. Why is this plate large?

Exercise:

Problem:

Some surgery is performed with high-voltage electricity passing from a metal scalpel through the tissue being cut. Considering the nature of electric fields at the surface of conductors, why would you expect most of the current to flow from the sharp edge of the scalpel? Do you think high- or low-frequency AC is used?

Exercise:

Problem:

Some devices often used in bathrooms, such as hairdryers, often have safety messages saying "Do not use when the bathtub or basin is full of water." Why is this so?

Exercise:

Problem:

We are often advised to not flick electric switches with wet hands, dry your hand first. We are also advised to never throw water on an electric fire. Why is this so?

Exercise:**Problem:**

Before working on a power transmission line, linemen will touch the line with the back of the hand as a final check that the voltage is zero. Why the back of the hand?

Exercise:**Problem:**

Why is the resistance of wet skin so much smaller than dry, and why do blood and other bodily fluids have low resistances?

Exercise:**Problem:**

Could a person on intravenous infusion (an IV) be microshock sensitive?

Exercise:**Problem:**

In view of the small currents that cause shock hazards and the larger currents that circuit breakers and fuses interrupt, how do they play a role in preventing shock hazards?

Problem Exercises**Exercise:****Problem:**

(a) How much power is dissipated in a short circuit of 240-V AC through a resistance of $0.250\ \Omega$? (b) What current flows?

Solution:

(a) 230 kW

(b) 960 A

Exercise:

Problem:

What voltage is involved in a 1.44-kW short circuit through a $0.100\text{ }\Omega$ resistance?

Exercise:

Problem:

Find the current through a person and identify the likely effect on her if she touches a 120-V AC source: (a) if she is standing on a rubber mat and offers a total resistance of $300\text{ k}\Omega$; (b) if she is standing barefoot on wet grass and has a resistance of only $4000\text{ k}\Omega$.

Solution:

(a) 0.400 mA, no effect

(b) 26.7 mA, muscular contraction for duration of the shock (can't let go)

Exercise:

Problem:

While taking a bath, a person touches the metal case of a radio. The path through the person to the drainpipe and ground has a resistance of $4000\text{ }\Omega$. What is the smallest voltage on the case of the radio that could cause ventricular fibrillation?

Exercise:

Problem:

Foolishly trying to fish a burning piece of bread from a toaster with a metal butter knife, a man comes into contact with 120-V AC. He does not even feel it since, luckily, he is wearing rubber-soled shoes. What is the minimum resistance of the path the current follows through the person?

Solution:

$$1.20 \times 10^5 \, \Omega$$

Exercise:**Problem:**

(a) During surgery, a current as small as $20.0 \, \mu\text{A}$ applied directly to the heart may cause ventricular fibrillation. If the resistance of the exposed heart is $300 \, \Omega$, what is the smallest voltage that poses this danger? (b) Does your answer imply that special electrical safety precautions are needed?

Exercise:**Problem:**

(a) What is the resistance of a 220-V AC short circuit that generates a peak power of 96.8 kW? (b) What would the average power be if the voltage was 120 V AC?

Solution:

(a) $1.00 \, \Omega$

(b) 14.4 kW

Exercise:

Problem:

A heart defibrillator passes 10.0 A through a patient's torso for 5.00 ms in an attempt to restore normal beating. (a) How much charge passed? (b) What voltage was applied if 500 J of energy was dissipated? (c) What was the path's resistance? (d) Find the temperature increase caused in the 8.00 kg of affected tissue.

Exercise:**Problem: Integrated Concepts**

A short circuit in a 120-V appliance cord has a $0.500\text{-}\Omega$ resistance. Calculate the temperature rise of the 2.00 g of surrounding materials, assuming their specific heat capacity is $0.200\text{ cal/g}\cdot^{\circ}\text{C}$ and that it takes 0.0500 s for a circuit breaker to interrupt the current. Is this likely to be damaging?

Solution:

Temperature increases 860°C . It is very likely to be damaging.

Exercise:**Problem: Construct Your Own Problem**

Consider a person working in an environment where electric currents might pass through her body. Construct a problem in which you calculate the resistance of insulation needed to protect the person from harm. Among the things to be considered are the voltage to which the person might be exposed, likely body resistance (dry, wet, ...), and acceptable currents (safe but sensed, safe and unfelt, ...).

Glossary

thermal hazard

a hazard in which electric current causes undesired thermal effects

shock hazard

when electric current passes through a person

short circuit

also known as a “short,” a low-resistance path between terminals of a voltage source

microshock sensitive

a condition in which a person’s skin resistance is bypassed, possibly by a medical procedure, rendering the person vulnerable to electrical shock at currents about 1/1000 the normally required level

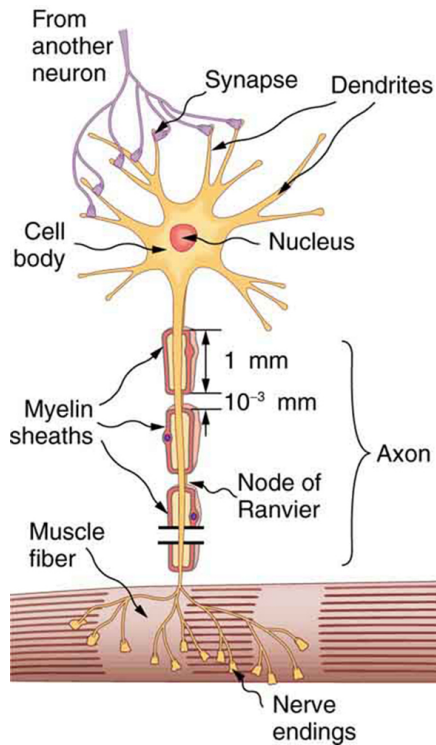
Nerve Conduction–Electrocardiograms

- Explain the process by which electric signals are transmitted along a neuron.
- Explain the effects myelin sheaths have on signal propagation.
- Explain what the features of an ECG signal indicate.

Nerve Conduction

Electric currents in the vastly complex system of billions of nerves in our body allow us to sense the world, control parts of our body, and think. These are representative of the three major functions of nerves. First, nerves carry messages from our sensory organs and others to the central nervous system, consisting of the brain and spinal cord. Second, nerves carry messages from the central nervous system to muscles and other organs. Third, nerves transmit and process signals within the central nervous system. The sheer number of nerve cells and the incredibly greater number of connections between them makes this system the subtle wonder that it is. **Nerve conduction** is a general term for electrical signals carried by nerve cells. It is one aspect of **bioelectricity**, or electrical effects in and created by biological systems.

Nerve cells, properly called *neurons*, look different from other cells—they have tendrils, some of them many centimeters long, connecting them with other cells. (See [\[link\]](#).) Signals arrive at the cell body across *synapses* or through *dendrites*, stimulating the neuron to generate its own signal, sent along its long *axon* to other nerve or muscle cells. Signals may arrive from many other locations and be transmitted to yet others, conditioning the synapses by use, giving the system its complexity and its ability to learn.

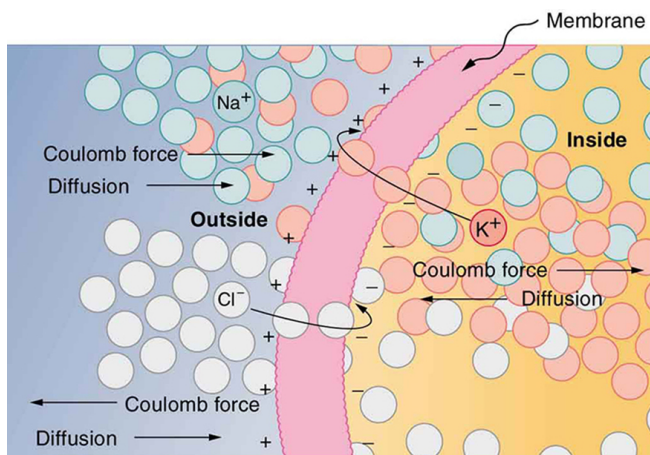


A neuron with its dendrites and long axon. Signals in the form of electric currents reach the cell body through dendrites and across synapses, stimulating the neuron to generate its own signal sent down the axon. The number of interconnections can be far greater than shown here.

The method by which these electric currents are generated and transmitted is more complex than the simple movement of free charges in a conductor,

but it can be understood with principles already discussed in this text. The most important of these are the Coulomb force and diffusion.

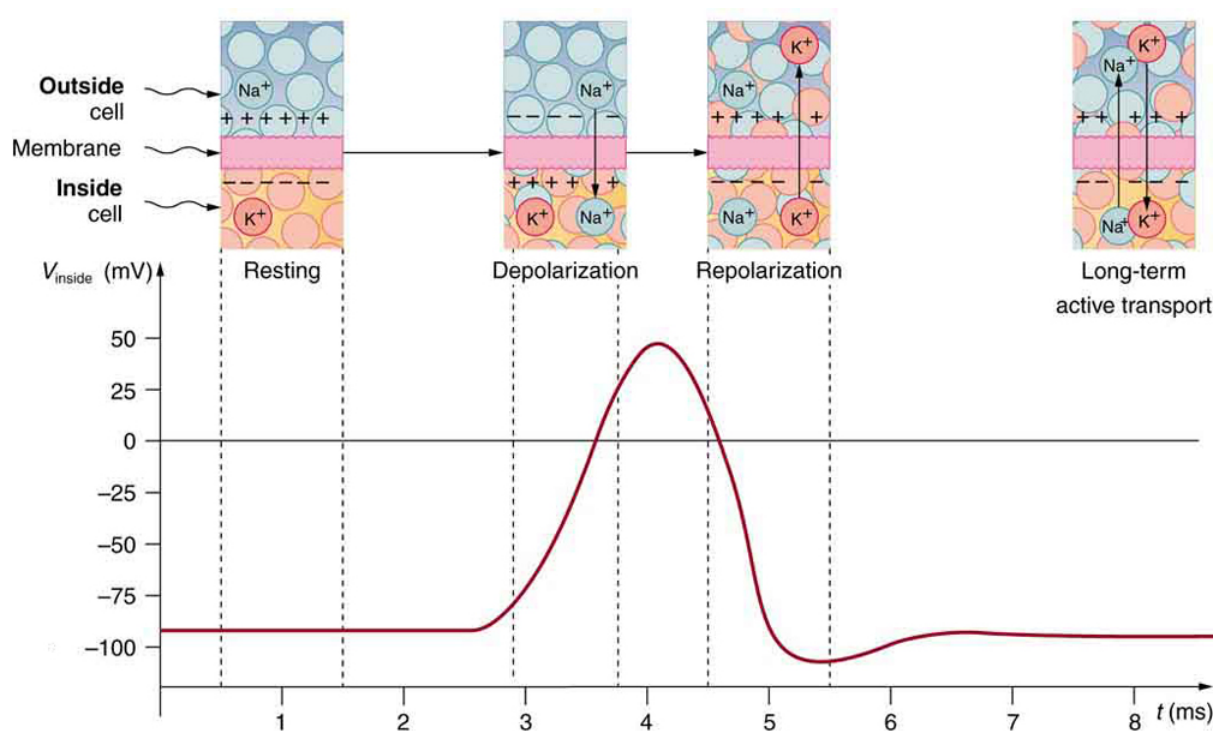
[\[link\]](#) illustrates how a voltage (potential difference) is created across the cell membrane of a neuron in its resting state. This thin membrane separates electrically neutral fluids having differing concentrations of ions, the most important varieties being Na^+ , K^+ , and Cl^- (these are sodium, potassium, and chlorine ions with single plus or minus charges as indicated). As discussed in [Molecular Transport Phenomena: Diffusion, Osmosis, and Related Processes](#), free ions will diffuse from a region of high concentration to one of low concentration. But the cell membrane is **semipermeable**, meaning that some ions may cross it while others cannot. In its resting state, the cell membrane is permeable to K^+ and Cl^- , and impermeable to Na^+ . Diffusion of K^+ and Cl^- thus creates the layers of positive and negative charge on the outside and inside of the membrane. The Coulomb force prevents the ions from diffusing across in their entirety. Once the charge layer has built up, the repulsion of like charges prevents more from moving across, and the attraction of unlike charges prevents more from leaving either side. The result is two layers of charge right on the membrane, with diffusion being balanced by the Coulomb force. A tiny fraction of the charges move across and the fluids remain neutral (other ions are present), while a separation of charge and a voltage have been created across the membrane.



The semipermeable membrane of a

cell has different concentrations of ions inside and out. Diffusion moves the K^+ and Cl^- ions in the direction shown, until the Coulomb force halts further transfer. This results in a layer of positive charge on the outside, a layer of negative charge on the inside, and thus a voltage across the cell membrane.

The membrane is normally impermeable to Na^+ .



An action potential is the pulse of voltage inside a nerve cell graphed here. It is caused by movements of ions across the cell membrane as shown. Depolarization occurs when a stimulus makes the membrane permeable to Na^+ ions. Repolarization follows as the membrane

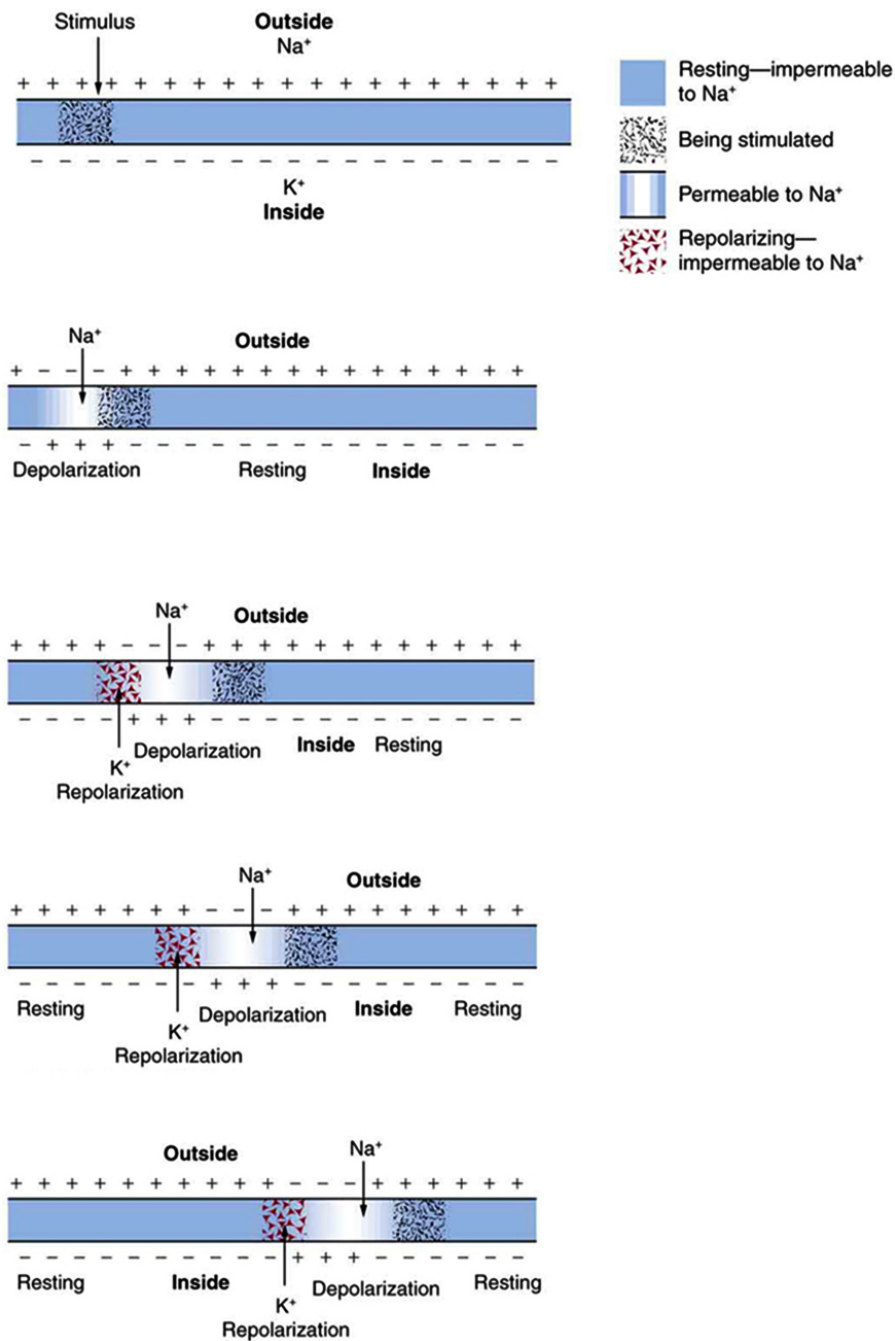
again becomes impermeable to Na^+ , and K^+ moves from high to low concentration. In the long term, active transport slowly maintains the concentration differences, but the cell may fire hundreds of times in rapid succession without seriously depleting them.

The separation of charge creates a potential difference of 70 to 90 mV across the cell membrane. While this is a small voltage, the resulting electric field ($E = V/d$) across the only 8-nm-thick membrane is immense (on the order of 11 MV/m!) and has fundamental effects on its structure and permeability. Now, if the exterior of a neuron is taken to be at 0 V, then the interior has a *resting potential* of about -90 mV. Such voltages are created across the membranes of almost all types of animal cells but are largest in nerve and muscle cells. In fact, fully 25% of the energy used by cells goes toward creating and maintaining these potentials.

Electric currents along the cell membrane are created by any stimulus that changes the membrane's permeability. The membrane thus temporarily becomes permeable to Na^+ , which then rushes in, driven both by diffusion and the Coulomb force. This inrush of Na^+ first neutralizes the inside membrane, or *depolarizes* it, and then makes it slightly positive. The depolarization causes the membrane to again become impermeable to Na^+ , and the movement of K^+ quickly returns the cell to its resting potential, or *repolarizes* it. This sequence of events results in a voltage pulse, called the *action potential*. (See [\[link\]](#).) Only small fractions of the ions move, so that the cell can fire many hundreds of times without depleting the excess concentrations of Na^+ and K^+ . Eventually, the cell must replenish these ions to maintain the concentration differences that create bioelectricity. This sodium-potassium pump is an example of *active transport*, wherein cell energy is used to move ions across membranes against diffusion gradients and the Coulomb force.

The action potential is a voltage pulse at one location on a cell membrane. How does it get transmitted along the cell membrane, and in particular down an axon, as a nerve impulse? The answer is that the changing voltage and electric fields affect the permeability of the adjacent cell membrane, so

that the same process takes place there. The adjacent membrane depolarizes, affecting the membrane further down, and so on, as illustrated in [\[link\]](#). Thus the action potential stimulated at one location triggers a *nerve impulse* that moves slowly (about 1 m/s) along the cell membrane.



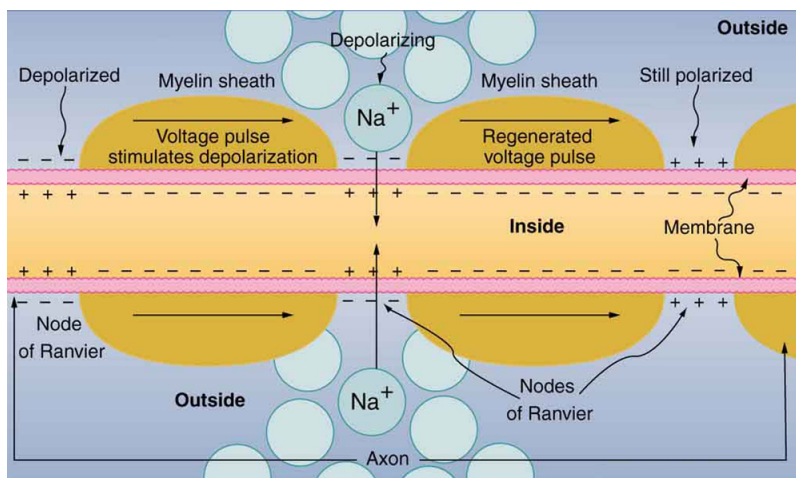
A nerve impulse is the propagation of an action potential along a cell membrane. A stimulus causes an action potential at one location, which changes the permeability of the adjacent membrane, causing an action potential there. This in turn affects the membrane further down, so that the action potential moves slowly (in electrical terms) along the cell membrane. Although the impulse is due to Na^+ and K^+ going across the membrane, it is equivalent to a wave of charge moving along the outside and inside of the membrane.

Some axons, like that in [\[link\]](#), are sheathed with *myelin*, consisting of fat-containing cells. [\[link\]](#) shows an enlarged view of an axon having myelin sheaths characteristically separated by unmyelinated gaps (called nodes of Ranvier). This arrangement gives the axon a number of interesting properties. Since myelin is an insulator, it prevents signals from jumping between adjacent nerves (cross talk). Additionally, the myelinated regions transmit electrical signals at a very high speed, as an ordinary conductor or resistor would. There is no action potential in the myelinated regions, so that no cell energy is used in them. There is an IR signal loss in the myelin, but the signal is regenerated in the gaps, where the voltage pulse triggers the action potential at full voltage. So a myelinated axon transmits a nerve impulse faster, with less energy consumption, and is better protected from cross talk than an unmyelinated one. Not all axons are myelinated, so that cross talk and slow signal transmission are a characteristic of the normal operation of these axons, another variable in the nervous system.

The degeneration or destruction of the myelin sheaths that surround the nerve fibers impairs signal transmission and can lead to numerous neurological effects. One of the most prominent of these diseases comes from the body's own immune system attacking the myelin in the central nervous system—multiple sclerosis. MS symptoms include fatigue, vision problems, weakness of arms and legs, loss of balance, and tingling or

numbness in one's extremities (neuropathy). It is more apt to strike younger adults, especially females. Causes might come from infection, environmental or geographic affects, or genetics. At the moment there is no known cure for MS.

Most animal cells can fire or create their own action potential. Muscle cells contract when they fire and are often induced to do so by a nerve impulse. In fact, nerve and muscle cells are physiologically similar, and there are even hybrid cells, such as in the heart, that have characteristics of both nerves and muscles. Some animals, like the infamous electric eel (see [\[link\]](#)), use muscles ganged so that their voltages add in order to create a shock great enough to stun prey.



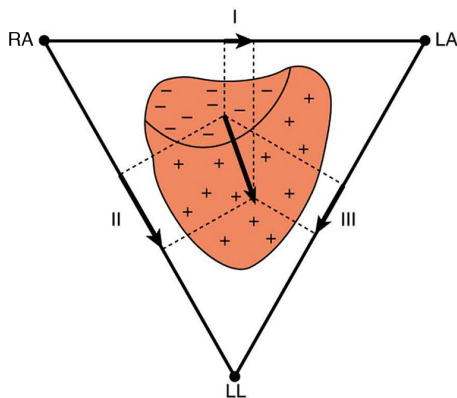
Propagation of a nerve impulse down a myelinated axon, from left to right. The signal travels very fast and without energy input in the myelinated regions, but it loses voltage. It is regenerated in the gaps. The signal moves faster than in unmyelinated axons and is insulated from signals in other nerves, limiting cross talk.



An electric eel flexes its muscles to create a voltage that stuns prey.
(credit: chrisbb, Flickr)

Electrocardiograms

Just as nerve impulses are transmitted by depolarization and repolarization of adjacent membrane, the depolarization that causes muscle contraction can also stimulate adjacent muscle cells to depolarize (fire) and contract. Thus, a depolarization wave can be sent across the heart, coordinating its rhythmic contractions and enabling it to perform its vital function of propelling blood through the circulatory system. [\[link\]](#) is a simplified graphic of a depolarization wave spreading across the heart from the *sinoarterial (SA) node*, the heart's natural pacemaker.



The outer surface of the heart changes from positive to negative during depolarization.

This wave of depolarization is spreading from the top of the heart and is represented by a vector pointing in the direction of the wave.

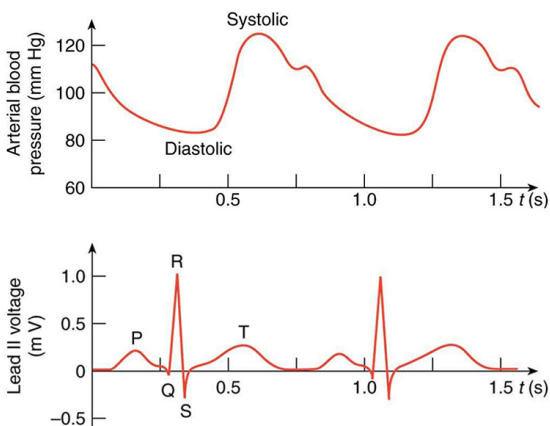
This vector is a voltage (potential difference) vector.

Three electrodes, labeled RA, LA, and LL, are placed on the patient. Each pair (called leads I, II, and III) measures a component of the depolarization vector and is graphed in an ECG.

An **electrocardiogram (ECG)** is a record of the voltages created by the wave of depolarization and subsequent repolarization in the heart. Voltages between pairs of electrodes placed on the chest are vector components of the voltage wave on the heart. Standard ECGs have 12 or more electrodes, but only three are shown in [\[link\]](#) for clarity. Decades ago, three-electrode ECGs were performed by placing electrodes on the left and right arms and the left leg. The voltage between the right arm and the left leg is called the *lead II potential* and is the most often graphed. We shall examine the lead II potential as an indicator of heart-muscle function and see that it is coordinated with arterial blood pressure as well.

Heart function and its four-chamber action are explored in [Viscosity and Laminar Flow; Poiseuille's Law](#). Basically, the right and left atria receive blood from the body and lungs, respectively, and pump the blood into the ventricles. The right and left ventricles, in turn, pump blood through the lungs and the rest of the body, respectively. Depolarization of the heart muscle causes it to contract. After contraction it is repolarized to ready it for the next beat. The ECG measures components of depolarization and repolarization of the heart muscle and can yield significant information on the functioning and malfunctioning of the heart.

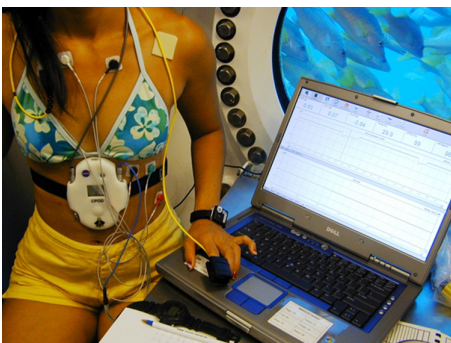
[\[link\]](#) shows an ECG of the lead II potential and a graph of the corresponding arterial blood pressure. The major features are labeled P, Q, R, S, and T. The *P wave* is generated by the depolarization and contraction of the atria as they pump blood into the ventricles. The *QRS complex* is created by the depolarization of the ventricles as they pump blood to the lungs and body. Since the shape of the heart and the path of the depolarization wave are not simple, the QRS complex has this typical shape and time span. The lead II QRS signal also masks the repolarization of the atria, which occur at the same time. Finally, the *T wave* is generated by the repolarization of the ventricles and is followed by the next P wave in the next heartbeat. Arterial blood pressure varies with each part of the heartbeat, with systolic (maximum) pressure occurring closely after the QRS complex, which signals contraction of the ventricles.



A lead II ECG with

corresponding arterial blood pressure. The QRS complex is created by the depolarization and contraction of the ventricles and is followed shortly by the maximum or systolic blood pressure. See text for further description.

Taken together, the 12 leads of a state-of-the-art ECG can yield a wealth of information about the heart. For example, regions of damaged heart tissue, called infarcts, reflect electrical waves and are apparent in one or more lead potentials. Subtle changes due to slight or gradual damage to the heart are most readily detected by comparing a recent ECG to an older one. This is particularly the case since individual heart shape, size, and orientation can cause variations in ECGs from one individual to another. ECG technology has advanced to the point where a portable ECG monitor with a liquid crystal instant display and a printer can be carried to patients' homes or used in emergency vehicles. See [\[link\]](#).



This NASA scientist and NEEMO 5 aquanaut's heart rate and other vital signs

are being recorded by
a portable device
while living in an
underwater habitat.
(credit: NASA, Life
Sciences Data Archive
at Johnson Space
Center, Houston,
Texas)

Note:

PhET Explorations: Neuron

Stimulate a neuron and monitor what happens. Pause, rewind, and move forward in time in order to observe the ions as they move across the neuron membrane.

https://phet.colorado.edu/sims/html/neuron/latest/neuron_en.html

Section Summary

- Electric potentials in neurons and other cells are created by ionic concentration differences across semipermeable membranes.
- Stimuli change the permeability and create action potentials that propagate along neurons.
- Myelin sheaths speed this process and reduce the needed energy input.
- This process in the heart can be measured with an electrocardiogram (ECG).

Conceptual Questions

Exercise:

Problem:

Note that in [\[link\]](#), both the concentration gradient and the Coulomb force tend to move Na^+ ions into the cell. What prevents this?

Exercise:**Problem:**

Define depolarization, repolarization, and the action potential.

Exercise:**Problem:**

Explain the properties of myelinated nerves in terms of the insulating properties of myelin.

Problems & Exercises**Exercise:****Problem: Integrated Concepts**

Use the ECG in [\[link\]](#) to determine the heart rate in beats per minute assuming a constant time between beats.

Solution:

80 beats/minute

Exercise:**Problem: Integrated Concepts**

(a) Referring to [\[link\]](#), find the time systolic pressure lags behind the middle of the QRS complex. (b) Discuss the reasons for the time lag.

Glossary

nerve conduction

the transport of electrical signals by nerve cells

bioelectricity

electrical effects in and created by biological systems

semipermeable

property of a membrane that allows only certain types of ions to cross it

electrocardiogram (ECG)

usually abbreviated ECG, a record of voltages created by depolarization and repolarization, especially in the heart

Introduction to Circuits and DC Instruments

class="introduction"

Electric
circuits in
a
computer
allow
large
amounts
of data to
be
quickly
and
accuratel
y
analyzed..
(credit:
Airman
1st Class
Mike
Meares,
United
States Air
Force)



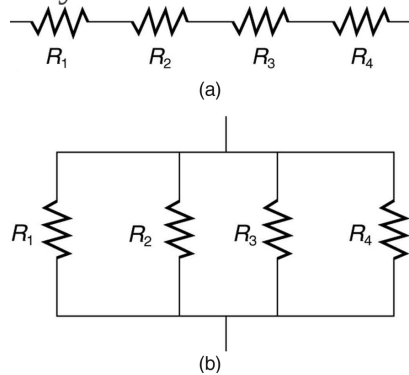
Electric circuits are commonplace. Some are simple, such as those in flashlights. Others, such as those used in supercomputers, are extremely complex.

This collection of modules takes the topic of electric circuits a step beyond simple circuits. When the circuit is purely resistive, everything in this module applies to both DC and AC. Matters become more complex when capacitance is involved. We do consider what happens when capacitors are connected to DC voltage sources, but the interaction of capacitors and other nonresistive devices with AC is left for a later chapter. Finally, a number of important DC instruments, such as meters that measure voltage and current, are covered in this chapter.

Resistors in Series and Parallel

- Draw a circuit with resistors in parallel and in series.
- Calculate the voltage drop of a current across a resistor using Ohm's law.
- Contrast the way total resistance is calculated for resistors in series and in parallel.
- Explain why total resistance of a parallel circuit is less than the smallest resistance of any of the resistors in that circuit.
- Calculate total resistance of a circuit that contains a mixture of resistors connected in series and in parallel.

Most circuits have more than one component, called a **resistor** that limits the flow of charge in the circuit. A measure of this limit on charge flow is called **resistance**. The simplest combinations of resistors are the series and parallel connections illustrated in [\[link\]](#). The total resistance of a combination of resistors depends on both their individual values and how they are connected.

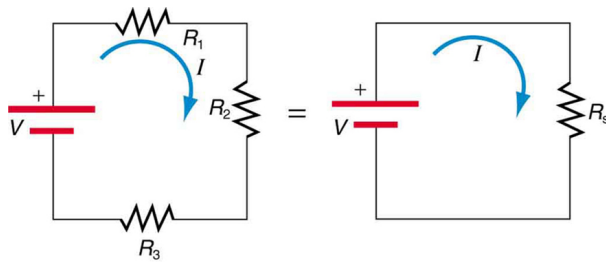


(a) A series connection of resistors. (b) A parallel connection of resistors.

Resistors in Series

When are resistors in **series**? Resistors are in series whenever the flow of charge, called the **current**, must flow through devices sequentially. For example, if current flows through a person holding a screwdriver and into the Earth, then R_1 in [\[link\]](#)(a) could be the resistance of the screwdriver's shaft, R_2 the resistance of its handle, R_3 the person's body resistance, and R_4 the resistance of her shoes.

[\[link\]](#) shows resistors in series connected to a **voltage** source. It seems reasonable that the total resistance is the sum of the individual resistances, considering that the current has to pass through each resistor in sequence. (This fact would be an advantage to a person wishing to avoid an electrical shock, who could reduce the current by wearing high-resistance rubber-soled shoes. It could be a disadvantage if one of the resistances were a faulty high-resistance cord to an appliance that would reduce the operating current.)



Three resistors connected in series to a battery (left) and the equivalent single or series resistance (right).

To verify that resistances in series do indeed add, let us consider the loss of electrical power, called a **voltage drop**, in each resistor in [\[link\]](#).

According to **Ohm's law**, the voltage drop, V , across a resistor when a current flows through it is calculated using the equation $V = IR$, where I equals the current in amps (A) and R is the resistance in ohms (Ω). Another

way to think of this is that V is the voltage necessary to make a current I flow through a resistance R .

So the voltage drop across R_1 is $V_1 = IR_1$, that across R_2 is $V_2 = IR_2$, and that across R_3 is $V_3 = IR_3$. The sum of these voltages equals the voltage output of the source; that is,

Equation:

$$V = V_1 + V_2 + V_3.$$

This equation is based on the conservation of energy and conservation of charge. Electrical potential energy can be described by the equation $PE = qV$, where q is the electric charge and V is the voltage. Thus the energy supplied by the source is qV , while that dissipated by the resistors is

Equation:

$$qV_1 + qV_2 + qV_3.$$

Note:

Connections: Conservation Laws

The derivations of the expressions for series and parallel resistance are based on the laws of conservation of energy and conservation of charge, which state that total charge and total energy are constant in any process. These two laws are directly involved in all electrical phenomena and will be invoked repeatedly to explain both specific effects and the general behavior of electricity.

These energies must be equal, because there is no other source and no other destination for energy in the circuit. Thus, $qV = qV_1 + qV_2 + qV_3$. The charge q cancels, yielding $V = V_1 + V_2 + V_3$, as stated. (Note that the same amount of charge passes through the battery and each resistor in a given amount of time, since there is no capacitance to store charge, there is no place for charge to leak, and charge is conserved.)

Now substituting the values for the individual voltages gives

Equation:

$$V = IR_1 + IR_2 + IR_3 = I(R_1 + R_2 + R_3).$$

Note that for the equivalent single series resistance R_s , we have

Equation:

$$V = IR_s.$$

This implies that the total or equivalent series resistance R_s of three resistors is $R_s = R_1 + R_2 + R_3$.

This logic is valid in general for any number of resistors in series; thus, the total resistance R_s of a series connection is

Equation:

$$R_s = R_1 + R_2 + R_3 + \dots,$$

as proposed. Since all of the current must pass through each resistor, it experiences the resistance of each, and resistances in series simply add up.

Example:

Calculating Resistance, Current, Voltage Drop, and Power

Dissipation: Analysis of a Series Circuit

Suppose the voltage output of the battery in [\[link\]](#) is 12.0 V, and the resistances are $R_1 = 1.00 \, \Omega$, $R_2 = 6.00 \, \Omega$, and $R_3 = 13.0 \, \Omega$. (a) What is the total resistance? (b) Find the current. (c) Calculate the voltage drop in each resistor, and show these add to equal the voltage output of the source. (d) Calculate the power dissipated by each resistor. (e) Find the power output of the source, and show that it equals the total power dissipated by the resistors.

Strategy and Solution for (a)

The total resistance is simply the sum of the individual resistances, as given by this equation:

Equation:

$$\begin{aligned}R_s &= R_1 + R_2 + R_3 \\&= 1.00\ \Omega + 6.00\ \Omega + 13.0\ \Omega \\&= 20.0\ \Omega.\end{aligned}$$

Strategy and Solution for (b)

The current is found using Ohm's law, $V = IR$. Entering the value of the applied voltage and the total resistance yields the current for the circuit:

Equation:

$$I = \frac{V}{R_s} = \frac{12.0\ \text{V}}{20.0\ \Omega} = 0.600\ \text{A}.$$

Strategy and Solution for (c)

The voltage—or IR drop—in a resistor is given by Ohm's law. Entering the current and the value of the first resistance yields

Equation:

$$V_1 = IR_1 = (0.600\ \text{A})(1.0\ \Omega) = 0.600\ \text{V}.$$

Similarly,

Equation:

$$V_2 = IR_2 = (0.600\ \text{A})(6.0\ \Omega) = 3.60\ \text{V}$$

and

Equation:

$$V_3 = IR_3 = (0.600\ \text{A})(13.0\ \Omega) = 7.80\ \text{V}.$$

Discussion for (c)

The three IR drops add to 12.0 V, as predicted:

Equation:

$$V_1 + V_2 + V_3 = (0.600 + 3.60 + 7.80)\ \text{V} = 12.0\ \text{V}.$$

Strategy and Solution for (d)

The easiest way to calculate power in watts (W) dissipated by a resistor in a DC circuit is to use **Joule's law**, $P = IV$, where P is electric power. In this case, each resistor has the same full current flowing through it. By substituting Ohm's law $V = IR$ into Joule's law, we get the power dissipated by the first resistor as

Equation:

$$P_1 = I^2 R_1 = (0.600 \text{ A})^2 (1.00 \Omega) = 0.360 \text{ W}.$$

Similarly,

Equation:

$$P_2 = I^2 R_2 = (0.600 \text{ A})^2 (6.00 \Omega) = 2.16 \text{ W}$$

and

Equation:

$$P_3 = I^2 R_3 = (0.600 \text{ A})^2 (13.0 \Omega) = 4.68 \text{ W}.$$

Discussion for (d)

Power can also be calculated using either $P = IV$ or $P = \frac{V^2}{R}$, where V is the voltage drop across the resistor (not the full voltage of the source). The same values will be obtained.

Strategy and Solution for (e)

The easiest way to calculate power output of the source is to use $P = IV$, where V is the source voltage. This gives

Equation:

$$P = (0.600 \text{ A})(12.0 \text{ V}) = 7.20 \text{ W}.$$

Discussion for (e)

Note, coincidentally, that the total power dissipated by the resistors is also 7.20 W, the same as the power put out by the source. That is,

Equation:

$$P_1 + P_2 + P_3 = (0.360 + 2.16 + 4.68) \text{ W} = 7.20 \text{ W}.$$

Power is energy per unit time (watts), and so conservation of energy requires the power output of the source to be equal to the total power dissipated by the resistors.

Note:

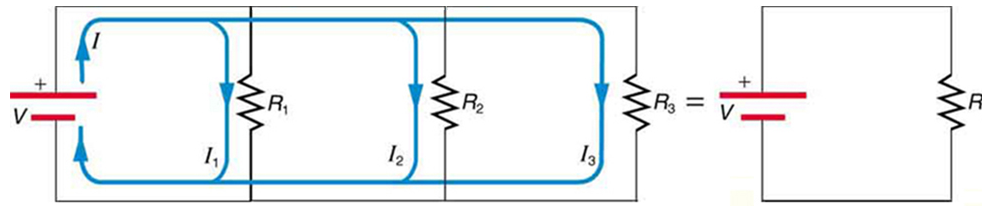
Major Features of Resistors in Series

1. Series resistances add: $R_s = R_1 + R_2 + R_3 + \dots$
2. The same current flows through each resistor in series.
3. Individual resistors in series do not get the total source voltage, but divide it.

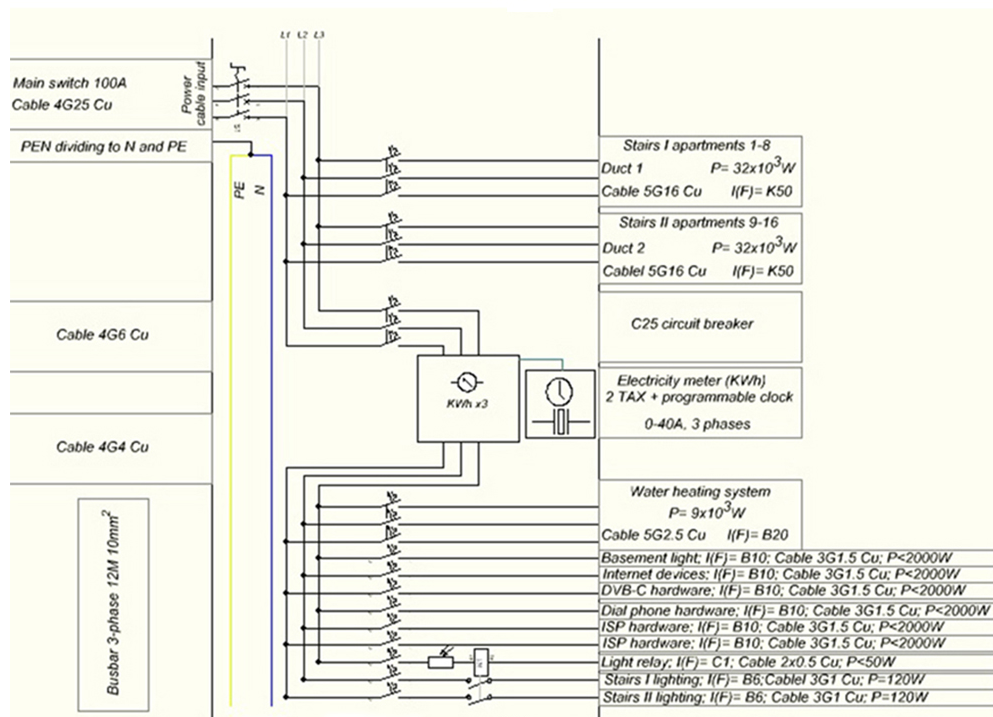
Resistors in Parallel

[\[link\]](#) shows resistors in **parallel**, wired to a voltage source. Resistors are in parallel when each resistor is connected directly to the voltage source by connecting wires having negligible resistance. Each resistor thus has the full voltage of the source applied to it.

Each resistor draws the same current it would if it alone were connected to the voltage source (provided the voltage source is not overloaded). For example, an automobile's headlights, radio, and so on, are wired in parallel, so that they utilize the full voltage of the source and can operate completely independently. The same is true in your house, or any building. (See [\[link\]](#) (b).)



(a)



(b)

(a) Three resistors connected in parallel to a battery and the equivalent single or parallel resistance. (b) Electrical power setup in a house. (credit: Dmitry G, Wikimedia Commons)

To find an expression for the equivalent parallel resistance R_p , let us consider the currents that flow and how they are related to resistance. Since each resistor in the circuit has the full voltage, the currents flowing through the individual resistors are $I_1 = \frac{V}{R_1}$, $I_2 = \frac{V}{R_2}$, and $I_3 = \frac{V}{R_3}$. Conservation of charge implies that the total current I produced by the source is the sum of these currents:

Equation:

$$I = I_1 + I_2 + I_3.$$

Substituting the expressions for the individual currents gives

Equation:

$$I = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3} = V \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right).$$

Note that Ohm's law for the equivalent single resistance gives

Equation:

$$I = \frac{V}{R_p} = V \left(\frac{1}{R_p} \right).$$

The terms inside the parentheses in the last two equations must be equal. Generalizing to any number of resistors, the total resistance R_p of a parallel connection is related to the individual resistances by

Equation:

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots$$

This relationship results in a total resistance R_p that is less than the smallest of the individual resistances. (This is seen in the next example.) When resistors are connected in parallel, more current flows from the source than would flow for any of them individually, and so the total resistance is lower.

Example:

Calculating Resistance, Current, Power Dissipation, and Power Output: Analysis of a Parallel Circuit

Let the voltage output of the battery and resistances in the parallel connection in [\[link\]](#) be the same as the previously considered series

connection: $V = 12.0 \text{ V}$, $R_1 = 1.00 \text{ } \Omega$, $R_2 = 6.00 \text{ } \Omega$, and $R_3 = 13.0 \text{ } \Omega$.

(a) What is the total resistance? (b) Find the total current. (c) Calculate the currents in each resistor, and show these add to equal the total current output of the source. (d) Calculate the power dissipated by each resistor. (e) Find the power output of the source, and show that it equals the total power dissipated by the resistors.

Strategy and Solution for (a)

The total resistance for a parallel combination of resistors is found using the equation below. Entering known values gives

Equation:

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} = \frac{1}{1.00 \text{ } \Omega} + \frac{1}{6.00 \text{ } \Omega} + \frac{1}{13.0 \text{ } \Omega}.$$

Thus,

Equation:

$$\frac{1}{R_p} = \frac{1.00}{\Omega} + \frac{0.1667}{\Omega} + \frac{0.07692}{\Omega} = \frac{1.2436}{\Omega}.$$

(Note that in these calculations, each intermediate answer is shown with an extra digit.)

We must invert this to find the total resistance R_p . This yields

Equation:

$$R_p = \frac{1}{1.2436} \Omega = 0.8041 \text{ } \Omega.$$

The total resistance with the correct number of significant digits is $R_p = 0.804 \text{ } \Omega$.

Discussion for (a)

R_p is, as predicted, less than the smallest individual resistance.

Strategy and Solution for (b)

The total current can be found from Ohm's law, substituting R_p for the total resistance. This gives

Equation:

$$I = \frac{V}{R_p} = \frac{12.0 \text{ V}}{0.8041 \Omega} = 14.92 \text{ A.}$$

Discussion for (b)

Current I for each device is much larger than for the same devices connected in series (see the previous example). A circuit with parallel connections has a smaller total resistance than the resistors connected in series.

Strategy and Solution for (c)

The individual currents are easily calculated from Ohm's law, since each resistor gets the full voltage. Thus,

Equation:

$$I_1 = \frac{V}{R_1} = \frac{12.0 \text{ V}}{1.00 \Omega} = 12.0 \text{ A.}$$

Similarly,

Equation:

$$I_2 = \frac{V}{R_2} = \frac{12.0 \text{ V}}{6.00 \Omega} = 2.00 \text{ A}$$

and

Equation:

$$I_3 = \frac{V}{R_3} = \frac{12.0 \text{ V}}{13.0 \Omega} = 0.92 \text{ A.}$$

Discussion for (c)

The total current is the sum of the individual currents:

Equation:

$$I_1 + I_2 + I_3 = 14.92 \text{ A.}$$

This is consistent with conservation of charge.

Strategy and Solution for (d)

The power dissipated by each resistor can be found using any of the equations relating power to current, voltage, and resistance, since all three

are known. Let us use $P = \frac{V^2}{R}$, since each resistor gets full voltage. Thus,
Equation:

$$P_1 = \frac{V^2}{R_1} = \frac{(12.0 \text{ V})^2}{1.00 \, \Omega} = 144 \text{ W}.$$

Similarly,
Equation:

$$P_2 = \frac{V^2}{R_2} = \frac{(12.0 \text{ V})^2}{6.00 \, \Omega} = 24.0 \text{ W}$$

and
Equation:

$$P_3 = \frac{V^2}{R_3} = \frac{(12.0 \text{ V})^2}{13.0 \, \Omega} = 11.1 \text{ W}.$$

Discussion for (d)

The power dissipated by each resistor is considerably higher in parallel than when connected in series to the same voltage source.

Strategy and Solution for (e)

The total power can also be calculated in several ways. Choosing $P = IV$, and entering the total current, yields

Equation:

$$P = IV = (14.92 \text{ A})(12.0 \text{ V}) = 179 \text{ W}.$$

Discussion for (e)

Total power dissipated by the resistors is also 179 W:

Equation:

$$P_1 + P_2 + P_3 = 144 \text{ W} + 24.0 \text{ W} + 11.1 \text{ W} = 179 \text{ W}.$$

This is consistent with the law of conservation of energy.

Overall Discussion

Note that both the currents and powers in parallel connections are greater than for the same devices in series.

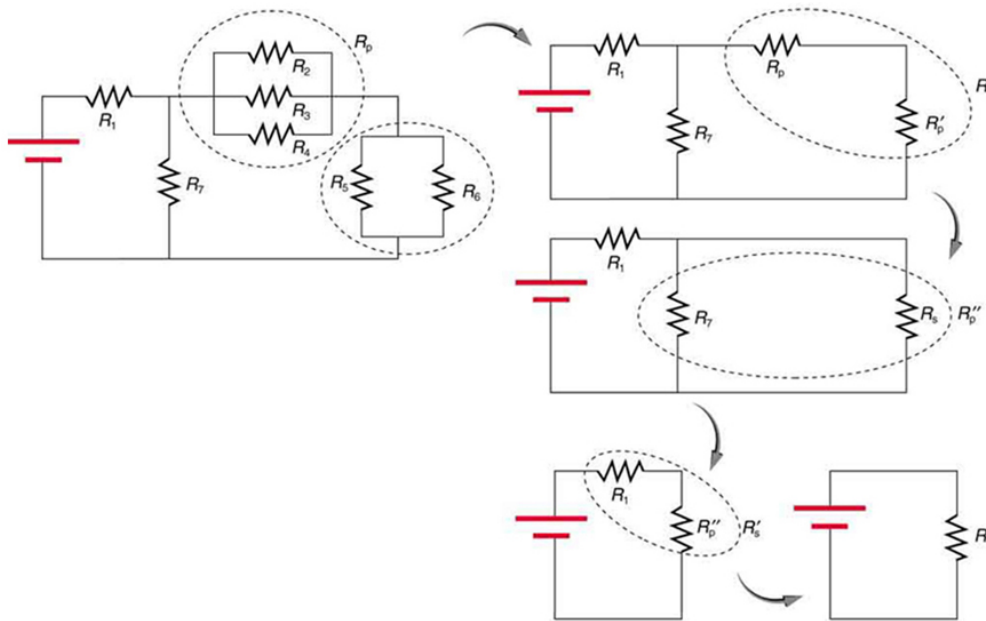
Note:**Major Features of Resistors in Parallel**

1. Parallel resistance is found from $\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots$, and it is smaller than any individual resistance in the combination.
2. Each resistor in parallel has the same full voltage of the source applied to it. (Power distribution systems most often use parallel connections to supply the myriad devices served with the same voltage and to allow them to operate independently.)
3. Parallel resistors do not each get the total current; they divide it.

Combinations of Series and Parallel

More complex connections of resistors are sometimes just combinations of series and parallel. These are commonly encountered, especially when wire resistance is considered. In that case, wire resistance is in series with other resistances that are in parallel.

Combinations of series and parallel can be reduced to a single equivalent resistance using the technique illustrated in [\[link\]](#). Various parts are identified as either series or parallel, reduced to their equivalents, and further reduced until a single resistance is left. The process is more time consuming than difficult.



This combination of seven resistors has both series and parallel parts. Each is identified and reduced to an equivalent resistance, and these are further reduced until a single equivalent resistance is reached.

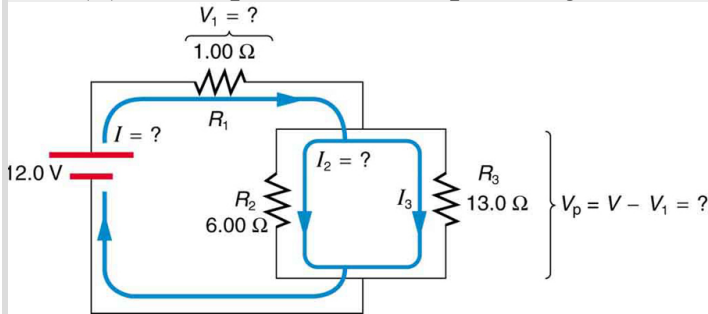
The simplest combination of series and parallel resistance, shown in [\[link\]](#), is also the most instructive, since it is found in many applications. For example, R_1 could be the resistance of wires from a car battery to its electrical devices, which are in parallel. R_2 and R_3 could be the starter motor and a passenger compartment light. We have previously assumed that wire resistance is negligible, but, when it is not, it has important effects, as the next example indicates.

Example:

Calculating Resistance, IR Drop, Current, and Power Dissipation: Combining Series and Parallel Circuits

[\[link\]](#) shows the resistors from the previous two examples wired in a different way—a combination of series and parallel. We can consider R_1 to be the resistance of wires leading to R_2 and R_3 . (a) Find the total

resistance. (b) What is the IR drop in R_1 ? (c) Find the current I_2 through R_2 . (d) What power is dissipated by R_2 ?



These three resistors are connected to a voltage source so that R_2 and R_3 are in parallel with one another and that combination is in series with R_1 .

Strategy and Solution for (a)

To find the total resistance, we note that R_2 and R_3 are in parallel and their combination R_p is in series with R_1 . Thus the total (equivalent) resistance of this combination is

Equation:

$$R_{\text{tot}} = R_1 + R_p.$$

First, we find R_p using the equation for resistors in parallel and entering known values:

Equation:

$$\frac{1}{R_p} = \frac{1}{R_2} + \frac{1}{R_3} = \frac{1}{6.00 \, \Omega} + \frac{1}{13.0 \, \Omega} = \frac{0.2436}{\Omega}.$$

Inverting gives

Equation:

$$R_p = \frac{1}{0.2436} \, \Omega = 4.11 \, \Omega.$$

So the total resistance is

Equation:

$$R_{\text{tot}} = R_1 + R_p = 1.00 \, \Omega + 4.11 \, \Omega = 5.11 \, \Omega.$$

Discussion for (a)

The total resistance of this combination is intermediate between the pure series and pure parallel values ($20.0 \, \Omega$ and $0.804 \, \Omega$, respectively) found for the same resistors in the two previous examples.

Strategy and Solution for (b)

To find the IR drop in R_1 , we note that the full current I flows through R_1 . Thus its IR drop is

Equation:

$$V_1 = IR_1.$$

We must find I before we can calculate V_1 . The total current I is found using Ohm's law for the circuit. That is,

Equation:

$$I = \frac{V}{R_{\text{tot}}} = \frac{12.0 \, \text{V}}{5.11 \, \Omega} = 2.35 \, \text{A}.$$

Entering this into the expression above, we get

Equation:

$$V_1 = IR_1 = (2.35 \, \text{A})(1.00 \, \Omega) = 2.35 \, \text{V}.$$

Discussion for (b)

The voltage applied to R_2 and R_3 is less than the total voltage by an amount V_1 . When wire resistance is large, it can significantly affect the operation of the devices represented by R_2 and R_3 .

Strategy and Solution for (c)

To find the current through R_2 , we must first find the voltage applied to it. We call this voltage V_p , because it is applied to a parallel combination of resistors. The voltage applied to both R_2 and R_3 is reduced by the amount V_1 , and so it is

Equation:

$$V_p = V - V_1 = 12.0 \text{ V} - 2.35 \text{ V} = 9.65 \text{ V}.$$

Now the current I_2 through resistance R_2 is found using Ohm's law:

Equation:

$$I_2 = \frac{V_p}{R_2} = \frac{9.65 \text{ V}}{6.00 \Omega} = 1.61 \text{ A}.$$

Discussion for (c)

The current is less than the 2.00 A that flowed through R_2 when it was connected in parallel to the battery in the previous parallel circuit example.

Strategy and Solution for (d)

The power dissipated by R_2 is given by

Equation:

$$P_2 = (I_2)^2 R_2 = (1.61 \text{ A})^2 (6.00 \Omega) = 15.5 \text{ W}.$$

Discussion for (d)

The power is less than the 24.0 W this resistor dissipated when connected in parallel to the 12.0-V source.

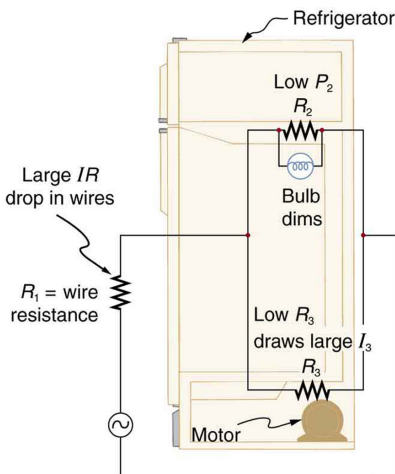
Practical Implications

One implication of this last example is that resistance in wires reduces the current and power delivered to a resistor. If wire resistance is relatively large, as in a worn (or a very long) extension cord, then this loss can be significant. If a large current is drawn, the IR drop in the wires can also be significant.

For example, when you are rummaging in the refrigerator and the motor comes on, the refrigerator light dims momentarily. Similarly, you can see the passenger compartment light dim when you start the engine of your car (although this may be due to resistance inside the battery itself).

What is happening in these high-current situations is illustrated in [\[link\]](#). The device represented by R_3 has a very low resistance, and so when it is

switched on, a large current flows. This increased current causes a larger IR drop in the wires represented by R_1 , reducing the voltage across the light bulb (which is R_2), which then dims noticeably.



Why do lights dim when a large appliance is switched on? The answer is that the large current the appliance motor draws causes a significant IR drop in the wires and reduces the voltage across the light.

Exercise:
Check Your Understanding

Problem:

Can any arbitrary combination of resistors be broken down into series and parallel combinations? See if you can draw a circuit diagram of resistors that cannot be broken down into combinations of series and parallel.

Solution:

No, there are many ways to connect resistors that are not combinations of series and parallel, including loops and junctions. In such cases Kirchhoff's rules, to be introduced in [Kirchhoff's Rules](#), will allow you to analyze the circuit.

Note:**Problem-Solving Strategies for Series and Parallel Resistors**

1. Draw a clear circuit diagram, labeling all resistors and voltage sources. This step includes a list of the knowns for the problem, since they are labeled in your circuit diagram.
2. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.
3. Determine whether resistors are in series, parallel, or a combination of both series and parallel. Examine the circuit diagram to make this assessment. Resistors are in series if the same current must pass sequentially through them.
4. Use the appropriate list of major features for series or parallel connections to solve for the unknowns. There is one list for series and another for parallel. If your problem has a combination of series and parallel, reduce it in steps by considering individual groups of series or parallel connections, as done in this module and the examples. Special note: When finding R_p , the reciprocal must be taken with care.
5. Check to see whether the answers are reasonable and consistent. Units and numerical results must be reasonable. Total series resistance

should be greater, whereas total parallel resistance should be smaller, for example. Power should be greater for the same devices in parallel compared with series, and so on.

Section Summary

- The total resistance of an electrical circuit with resistors wired in a series is the sum of the individual resistances:
 $R_s = R_1 + R_2 + R_3 + \dots$
- Each resistor in a series circuit has the same amount of current flowing through it.
- The voltage drop, or power dissipation, across each individual resistor in a series is different, and their combined total adds up to the power source input.
- The total resistance of an electrical circuit with resistors wired in parallel is less than the lowest resistance of any of the components and can be determined using the formula:

Equation:

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots$$

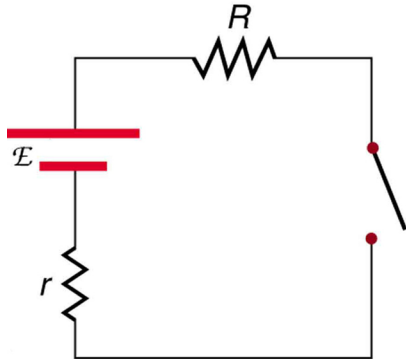
- Each resistor in a parallel circuit has the same full voltage of the source applied to it.
- The current flowing through each resistor in a parallel circuit is different, depending on the resistance.
- If a more complex connection of resistors is a combination of series and parallel, it can be reduced to a single equivalent resistance by identifying its various parts as series or parallel, reducing each to its equivalent, and continuing until a single resistance is eventually reached.

Conceptual Questions

Exercise:

Problem:

A switch has a variable resistance that is nearly zero when closed and extremely large when open, and it is placed in series with the device it controls. Explain the effect the switch in [\[link\]](#) has on current when open and when closed.



A switch is ordinarily in series with a resistance and voltage source. Ideally, the switch has nearly zero resistance when closed but has an extremely large resistance when open. (Note that in this diagram, the script E represents the voltage (or electromotive force) of the battery.)

Exercise:

Problem: What is the voltage across the open switch in [\[link\]](#)?

Exercise:

Problem:

There is a voltage across an open switch, such as in [\[link\]](#). Why, then, is the power dissipated by the open switch small?

Exercise:

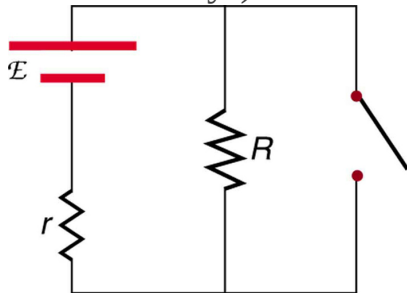
Problem:

Why is the power dissipated by a closed switch, such as in [\[link\]](#), small?

Exercise:

Problem:

A student in a physics lab mistakenly wired a light bulb, battery, and switch as shown in [\[link\]](#). Explain why the bulb is on when the switch is open, and off when the switch is closed. (Do not try this—it is hard on the battery!)



A wiring mistake put this switch in parallel with the device represented by R . (Note that in this diagram, the script E represents the voltage (or

electromotive
force) of the
battery.)

Exercise:

Problem:

Knowing that the severity of a shock depends on the magnitude of the current through your body, would you prefer to be in series or parallel with a resistance, such as the heating element of a toaster, if shocked by it? Explain.

Exercise:

Problem:

Would your headlights dim when you start your car's engine if the wires in your automobile were superconductors? (Do not neglect the battery's internal resistance.) Explain.

Exercise:

Problem:

Some strings of holiday lights are wired in series to save wiring costs. An old version utilized bulbs that break the electrical connection, like an open switch, when they burn out. If one such bulb burns out, what happens to the others? If such a string operates on 120 V and has 40 identical bulbs, what is the normal operating voltage of each? Newer versions use bulbs that short circuit, like a closed switch, when they burn out. If one such bulb burns out, what happens to the others? If such a string operates on 120 V and has 39 remaining identical bulbs, what is then the operating voltage of each?

Exercise:

Problem:

If two household lightbulbs rated 60 W and 100 W are connected in series to household power, which will be brighter? Explain.

Exercise:**Problem:**

Suppose you are doing a physics lab that asks you to put a resistor into a circuit, but all the resistors supplied have a larger resistance than the requested value. How would you connect the available resistances to attempt to get the smaller value asked for?

Exercise:**Problem:**

Before World War II, some radios got power through a “resistance cord” that had a significant resistance. Such a resistance cord reduces the voltage to a desired level for the radio’s tubes and the like, and it saves the expense of a transformer. Explain why resistance cords become warm and waste energy when the radio is on.

Exercise:**Problem:**

Some light bulbs have three power settings (not including zero), obtained from multiple filaments that are individually switched and wired in parallel. What is the minimum number of filaments needed for three power settings?

Problem Exercises

Note: Data taken from figures can be assumed to be accurate to three significant digits.

Exercise:

Problem:

- (a) What is the resistance of ten $275\text{-}\Omega$ resistors connected in series?
(b) In parallel?
-

Solution:

- (a) $2.75\text{ k}\Omega$
(b) $27.5\text{ }\Omega$

Exercise:**Problem:**

- (a) What is the resistance of a $1.00 \times 10^2\text{-}\Omega$, a $2.50\text{-k}\Omega$, and a $4.00\text{-k}\Omega$ resistor connected in series? (b) In parallel?

Exercise:**Problem:**

What are the largest and smallest resistances you can obtain by connecting a $36.0\text{-}\Omega$, a $50.0\text{-}\Omega$, and a $700\text{-}\Omega$ resistor together?

Solution:

- (a) $786\text{ }\Omega$
(b) $20.3\text{ }\Omega$

Exercise:**Problem:**

An 1800-W toaster, a 1400-W electric frying pan, and a 75-W lamp are plugged into the same outlet in a 15-A , 120-V circuit. (The three devices are in parallel when plugged into the same socket.). (a) What current is drawn by each device? (b) Will this combination blow the 15-A fuse?

Exercise:

Problem:

Your car's 30.0-W headlight and 2.40-kW starter are ordinarily connected in parallel in a 12.0-V system. What power would one headlight and the starter consume if connected in series to a 12.0-V battery? (Neglect any other resistance in the circuit and any change in resistance in the two devices.)

Solution:

29.6 W

Exercise:**Problem:**

(a) Given a 48.0-V battery and $24.0\text{-}\Omega$ and $96.0\text{-}\Omega$ resistors, find the current and power for each when connected in series. (b) Repeat when the resistances are in parallel.

Exercise:**Problem:**

Referring to the example combining series and parallel circuits and [\[link\]](#), calculate I_3 in the following two different ways: (a) from the known values of I and I_2 ; (b) using Ohm's law for R_3 . In both parts explicitly show how you follow the steps in the [Problem-Solving Strategies for Series and Parallel Resistors](#).

Solution:

(a) 0.74 A

(b) 0.742 A

Exercise:

Problem:

Referring to [\[link\]](#): (a) Calculate P_3 and note how it compares with P_3 found in the first two example problems in this module. (b) Find the total power supplied by the source and compare it with the sum of the powers dissipated by the resistors.

Exercise:**Problem:**

Refer to [\[link\]](#) and the discussion of lights dimming when a heavy appliance comes on. (a) Given the voltage source is 120 V, the wire resistance is $0.400\ \Omega$, and the bulb is nominally 75.0 W, what power will the bulb dissipate if a total of 15.0 A passes through the wires when the motor comes on? Assume negligible change in bulb resistance. (b) What power is consumed by the motor?

Solution:

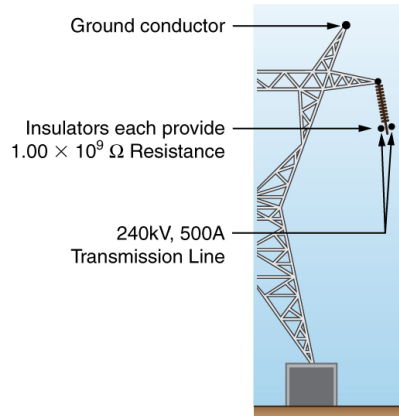
(a) 60.8 W

(b) 3.18 kW

Exercise:**Problem:**

A 240-kV power transmission line carrying $5.00 \times 10^2\ \text{A}$ is hung from grounded metal towers by ceramic insulators, each having a $1.00 \times 10^9\ \Omega$ resistance. [\[link\]](#). (a) What is the resistance to ground of 100 of these insulators? (b) Calculate the power dissipated by 100 of them. (c) What fraction of the power carried by the line is this?

Explicitly show how you follow the steps in the [Problem-Solving Strategies for Series and Parallel Resistors](#).



High-voltage (240-kV) transmission line carrying $5.00 \times 10^2 \text{ A}$ is hung from a grounded metal transmission tower. The row of ceramic insulators provide $1.00 \times 10^9 \Omega$ of resistance each.

Exercise:

Problem:

Show that if two resistors R_1 and R_2 are combined and one is much greater than the other ($R_1 \gg R_2$): (a) Their series resistance is very nearly equal to the greater resistance R_1 . (b) Their parallel resistance is very nearly equal to smaller resistance R_2 .

Solution:

$$R_s = R_1 + R_2$$

(a) $\Rightarrow R_s \approx R_1 (R_1 \gg R_2)$

$$(b) \frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{R_1 + R_2}{R_1 R_2},$$

so that

$$R_p = \frac{R_1 R_2}{R_1 + R_2} \approx \frac{R_1 R_2}{R_1} = R_2 (R_1 \gg R_2).$$

Exercise:

Problem: Unreasonable Results

Two resistors, one having a resistance of $145 \, \Omega$, are connected in parallel to produce a total resistance of $150 \, \Omega$. (a) What is the value of the second resistance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: Unreasonable Results

Two resistors, one having a resistance of $900 \, \text{k}\Omega$, are connected in series to produce a total resistance of $0.500 \, \text{M}\Omega$. (a) What is the value of the second resistance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $-400 \, \text{k}\Omega$

(b) Resistance cannot be negative.

(c) Series resistance is said to be less than one of the resistors, but it must be greater than any of the resistors.

Glossary

series

a sequence of resistors or other components wired into a circuit one after the other

resistor

a component that provides resistance to the current flowing through an electrical circuit

resistance

causing a loss of electrical power in a circuit

Ohm's law

the relationship between current, voltage, and resistance within an electrical circuit: $V = IR$

voltage

the electrical potential energy per unit charge; electric pressure created by a power source, such as a battery

voltage drop

the loss of electrical power as a current travels through a resistor, wire or other component

current

the flow of charge through an electric circuit past a given point of measurement

Joule's law

the relationship between potential electrical power, voltage, and resistance in an electrical circuit, given by: $P_e = IV$

parallel

the wiring of resistors or other components in an electrical circuit such that each component receives an equal voltage from the power source; often pictured in a ladder-shaped diagram, with each component on a rung of the ladder

Electromotive Force: Terminal Voltage

- Compare and contrast the voltage and the electromagnetic force of an electric power source.
- Describe what happens to the terminal voltage, current, and power delivered to a load as internal resistance of the voltage source increases (due to aging of batteries, for example).
- Explain why it is beneficial to use more than one voltage source connected in parallel.

When you forget to turn off your car lights, they slowly dim as the battery runs down. Why don't they simply blink off when the battery's energy is gone? Their gradual dimming implies that battery output voltage decreases as the battery is depleted.

Furthermore, if you connect an excessive number of 12-V lights in parallel to a car battery, they will be dim even when the battery is fresh and even if the wires to the lights have very low resistance. This implies that the battery's output voltage is reduced by the overload.

The reason for the decrease in output voltage for depleted or overloaded batteries is that all voltage sources have two fundamental parts—a source of electrical energy and an **internal resistance**. Let us examine both.

Electromotive Force

You can think of many different types of voltage sources. Batteries themselves come in many varieties. There are many types of mechanical/electrical generators, driven by many different energy sources, ranging from nuclear to wind. Solar cells create voltages directly from light, while thermoelectric devices create voltage from temperature differences.

A few voltage sources are shown in [\[link\]](#). All such devices create a **potential difference** and can supply current if connected to a resistance. On the small scale, the potential difference creates an electric field that exerts force on charges, causing current. We thus use the name **electromotive force**, abbreviated emf.

Emf is not a force at all; it is a special type of potential difference. To be precise, the electromotive force (emf) is the potential difference of a source when no current is flowing. Units of emf are volts.



A variety of voltage sources (clockwise from top left): the Brazos Wind Farm in Fluvanna, Texas (credit: Leaflet, Wikimedia Commons); the Krasnoyarsk Dam in Russia (credit: Alex Polezhaev); a solar farm (credit: U.S. Department of Energy); and a group of nickel metal hydride batteries (credit: Tiaa Monto). The voltage output of each depends on its construction and load, and equals emf only if there is no load.

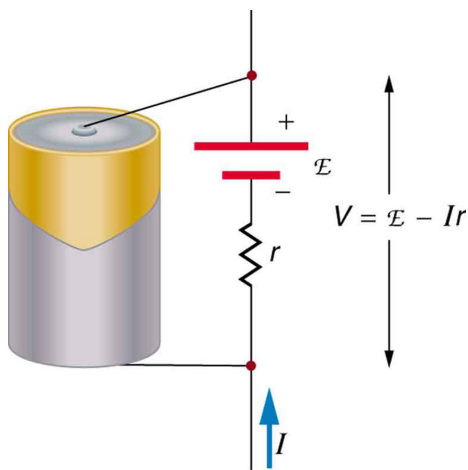
Electromotive force is directly related to the source of potential difference, such as the particular combination of chemicals in a battery. However, emf

differs from the voltage output of the device when current flows. The voltage across the terminals of a battery, for example, is less than the emf when the battery supplies current, and it declines further as the battery is depleted or loaded down. However, if the device's output voltage can be measured without drawing current, then output voltage will equal emf (even for a very depleted battery).

Internal Resistance

As noted before, a 12-V truck battery is physically larger, contains more charge and energy, and can deliver a larger current than a 12-V motorcycle battery. Both are lead-acid batteries with identical emf, but, because of its size, the truck battery has a smaller internal resistance r . Internal resistance is the inherent resistance to the flow of current within the source itself.

[\[link\]](#) is a schematic representation of the two fundamental parts of any voltage source. The emf (represented by a script \mathcal{E} in the figure) and internal resistance r are in series. The smaller the internal resistance for a given emf, the more current and the more power the source can supply.



Any voltage source (in this case, a carbon-zinc dry cell) has an emf related to its source of

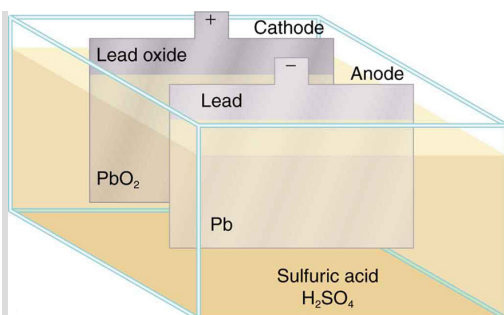
potential difference,
and an internal
resistance r related to
its construction. (Note
that the script E stands
for emf.). Also shown
are the output
terminals across which
the terminal voltage V
is measured. Since
 $V = \text{emf} - Ir$,
terminal voltage equals
emf only if there is no
current flowing.

The internal resistance r can behave in complex ways. As noted, r increases as a battery is depleted. But internal resistance may also depend on the magnitude and direction of the current through a voltage source, its temperature, and even its history. The internal resistance of rechargeable nickel-cadmium cells, for example, depends on how many times and how deeply they have been depleted.

Note:

Things Great and Small: The Submicroscopic Origin of Battery Potential
Various types of batteries are available, with emfs determined by the combination of chemicals involved. We can view this as a molecular reaction (what much of chemistry is about) that separates charge.

The lead-acid battery used in cars and other vehicles is one of the most common types. A single cell (one of six) of this battery is seen in [\[link\]](#). The cathode (positive) terminal of the cell is connected to a lead oxide plate, while the anode (negative) terminal is connected to a lead plate. Both plates are immersed in sulfuric acid, the electrolyte for the system.

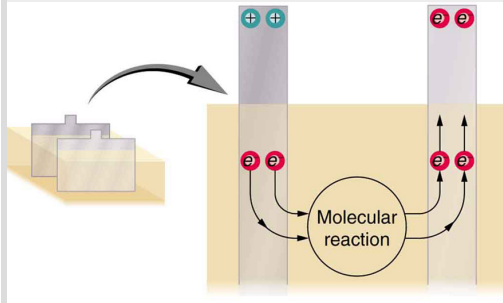


Artist's conception of a lead-acid cell. Chemical reactions in a lead-acid cell separate charge, sending negative charge to the anode, which is connected to the lead plates. The lead oxide plates are connected to the positive or cathode terminal of the cell. Sulfuric acid conducts the charge as well as participating in the chemical reaction.

The details of the chemical reaction are left to the reader to pursue in a chemistry text, but their results at the molecular level help explain the potential created by the battery. [\[link\]](#) shows the result of a single chemical reaction. Two electrons are placed on the anode, making it negative, provided that the cathode supplied two electrons. This leaves the cathode positively charged, because it has lost two electrons. In short, a separation of charge has been driven by a chemical reaction.

Note that the reaction will not take place unless there is a complete circuit to allow two electrons to be supplied to the cathode. Under many circumstances, these electrons come from the anode, flow through a resistance, and return to the cathode. Note also that since the chemical

reactions involve substances with resistance, it is not possible to create the emf without an internal resistance.



Artist's conception of two electrons being forced onto the anode of a cell and two electrons being removed from the cathode of the cell. The chemical reaction in a lead-acid battery places two electrons on the anode and removes two from the cathode. It requires a closed circuit to proceed, since the two electrons must be supplied to the cathode.

Why are the chemicals able to produce a unique potential difference? Quantum mechanical descriptions of molecules, which take into account the types of atoms and numbers of electrons in them, are able to predict the energy states they can have and the energies of reactions between them.

In the case of a lead-acid battery, an energy of 2 eV is given to each electron sent to the anode. Voltage is defined as the electrical potential

energy divided by charge: $V = \frac{P_E}{q}$. An electron volt is the energy given to a single electron by a voltage of 1 V. So the voltage here is 2 V, since 2 eV is given to each electron. It is the energy produced in each molecular reaction that produces the voltage. A different reaction produces a different energy and, hence, a different voltage.

Terminal Voltage

The voltage output of a device is measured across its terminals and, thus, is called its **terminal voltage** V . Terminal voltage is given by

Equation:

$$V = \text{emf} - Ir,$$

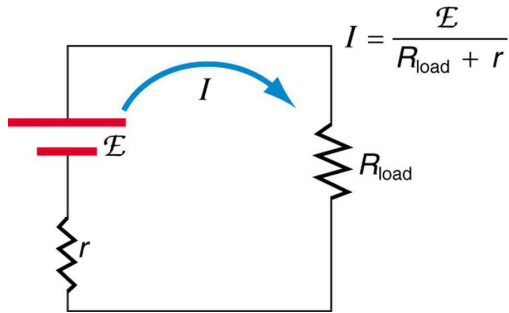
where r is the internal resistance and I is the current flowing at the time of the measurement.

I is positive if current flows away from the positive terminal, as shown in [\[link\]](#). You can see that the larger the current, the smaller the terminal voltage. And it is likewise true that the larger the internal resistance, the smaller the terminal voltage.

Suppose a load resistance R_{load} is connected to a voltage source, as in [\[link\]](#). Since the resistances are in series, the total resistance in the circuit is $R_{\text{load}} + r$. Thus the current is given by Ohm's law to be

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r}.$$



Schematic of a voltage source and its load R_{load} .

Since the internal resistance r is in series with the load, it can significantly affect the terminal voltage and current delivered to the load. (Note that the script E stands for emf.)

We see from this expression that the smaller the internal resistance r , the greater the current the voltage source supplies to its load R_{load} . As batteries are depleted, r increases. If r becomes a significant fraction of the load resistance, then the current is significantly reduced, as the following example illustrates.

Example:

Calculating Terminal Voltage, Power Dissipation, Current, and Resistance: Terminal Voltage and Load

A certain battery has a 12.0-V emf and an internal resistance of 0.100 Ω .

(a) Calculate its terminal voltage when connected to a 10.0- Ω load. (b) What is the terminal voltage when connected to a 0.500- Ω load? (c) What power does the 0.500- Ω load dissipate? (d) If the internal resistance grows

to $0.500\ \Omega$, find the current, terminal voltage, and power dissipated by a $0.500\text{-}\Omega$ load.

Strategy

The analysis above gave an expression for current when internal resistance is taken into account. Once the current is found, the terminal voltage can be calculated using the equation $V = \text{emf} - Ir$. Once current is found, the power dissipated by a resistor can also be found.

Solution for (a)

Entering the given values for the emf, load resistance, and internal resistance into the expression above yields

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\ \text{V}}{10.1\ \Omega} = 1.188\ \text{A}.$$

Enter the known values into the equation $V = \text{emf} - Ir$ to get the terminal voltage:

Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\ \text{V} - (1.188\ \text{A})(0.100\ \Omega) \\ &= 11.9\ \text{V}. \end{aligned}$$

Discussion for (a)

The terminal voltage here is only slightly lower than the emf, implying that $10.0\ \Omega$ is a light load for this particular battery.

Solution for (b)

Similarly, with $R_{\text{load}} = 0.500\ \Omega$, the current is

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\ \text{V}}{0.600\ \Omega} = 20.0\ \text{A}.$$

The terminal voltage is now

Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\ \text{V} - (20.0\ \text{A})(0.100\ \Omega) \\ &= 10.0\ \text{V}. \end{aligned}$$

Discussion for (b)

This terminal voltage exhibits a more significant reduction compared with emf, implying $0.500\ \Omega$ is a heavy load for this battery.

Solution for (c)

The power dissipated by the $0.500\text{-}\Omega$ load can be found using the formula $P = I^2 R$. Entering the known values gives

Equation:

$$P_{\text{load}} = I^2 R_{\text{load}} = (20.0\text{ A})^2 (0.500\ \Omega) = 2.00 \times 10^2\text{ W}.$$

Discussion for (c)

Note that this power can also be obtained using the expressions $\frac{V^2}{R}$ or IV , where V is the terminal voltage (10.0 V in this case).

Solution for (d)

Here the internal resistance has increased, perhaps due to the depletion of the battery, to the point where it is as great as the load resistance. As before, we first find the current by entering the known values into the expression, yielding

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\text{ V}}{1.00\ \Omega} = 12.0\text{ A}.$$

Now the terminal voltage is

Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\text{ V} - (12.0\text{ A})(0.500\ \Omega) \\ &= 6.00\text{ V}, \end{aligned}$$

and the power dissipated by the load is

Equation:

$$P_{\text{load}} = I^2 R_{\text{load}} = (12.0\text{ A})^2 (0.500\ \Omega) = 72.0\text{ W}.$$

Discussion for (d)

We see that the increased internal resistance has significantly decreased terminal voltage, current, and power delivered to a load.

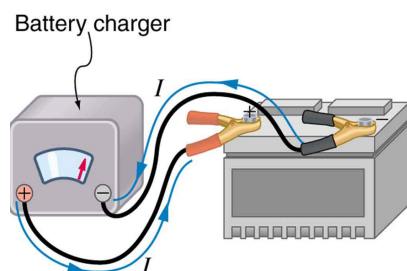
Battery testers, such as those in [\[link\]](#), use small load resistors to intentionally draw current to determine whether the terminal voltage drops below an acceptable level. They really test the internal resistance of the battery. If internal resistance is high, the battery is weak, as evidenced by its low terminal voltage.



These two battery testers measure terminal voltage under a load to determine the condition of a battery. The large device is being used by a U.S. Navy electronics technician to test large batteries aboard the aircraft carrier USS *Nimitz* and has a small resistance that can dissipate large amounts of power. (credit: U.S. Navy photo by Photographer's Mate Airman Jason A. Johnston) The small device is used on small batteries and has a digital display to indicate the acceptability of their terminal voltage. (credit: Keith Williamson)

Some batteries can be recharged by passing a current through them in the direction opposite to the current they supply to a resistance. This is done routinely in cars and batteries for small electrical appliances and electronic devices, and is represented pictorially in [\[link\]](#). The voltage output of the battery charger must be greater than the emf of the battery to reverse current

through it. This will cause the terminal voltage of the battery to be greater than the emf, since $V = \text{emf} - Ir$, and I is now negative.



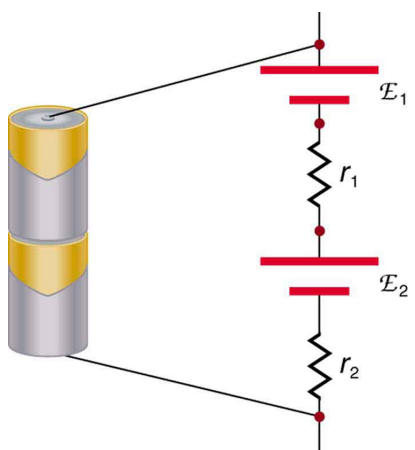
A car battery charger reverses the normal direction of current through a battery, reversing its chemical reaction and replenishing its chemical potential.

Multiple Voltage Sources

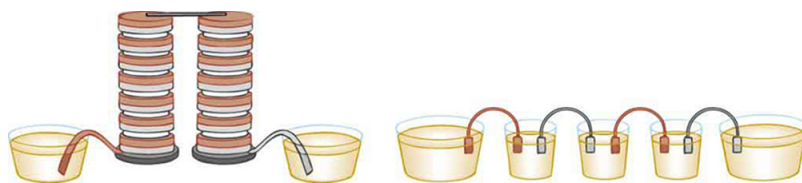
There are two voltage sources when a battery charger is used. Voltage sources connected in series are relatively simple. When voltage sources are in series, their internal resistances add and their emfs add algebraically. (See [\[link\]](#).) Series connections of voltage sources are common—for example, in flashlights, toys, and other appliances. Usually, the cells are in series in order to produce a larger total emf.

But if the cells oppose one another, such as when one is put into an appliance backward, the total emf is less, since it is the algebraic sum of the individual emfs.

A battery is a multiple connection of voltaic cells, as shown in [\[link\]](#). The disadvantage of series connections of cells is that their internal resistances add. One of the authors once owned a 1957 MGA that had two 6-V batteries in series, rather than a single 12-V battery. This arrangement produced a large internal resistance that caused him many problems in starting the engine.

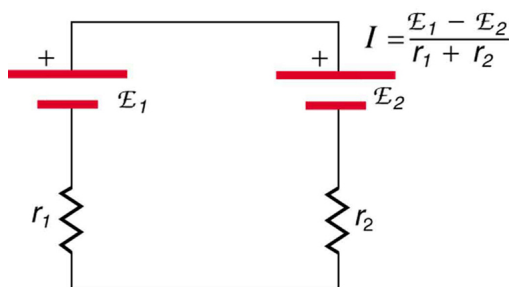


A series connection
of two voltage
sources. The emfs
(each labeled with
a script E) and
internal resistances
add, giving a total
emf of
 $\text{emf}_1 + \text{emf}_2$ and a
total internal
resistance of
 $r_1 + r_2$.



Batteries are multiple connections of individual cells, as shown in this modern rendition of an old print. Single cells, such as AA or C cells, are commonly called batteries, although this is technically incorrect.

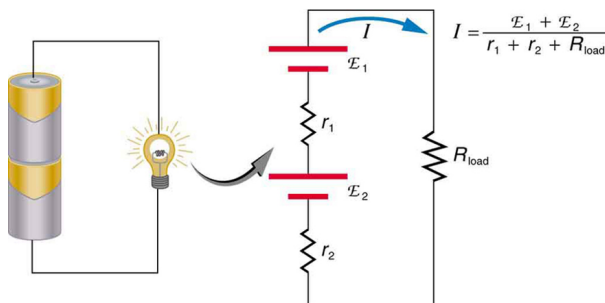
If the *series* connection of two voltage sources is made into a complete circuit with the emfs in opposition, then a current of magnitude $I = \frac{(\text{emf}_1 - \text{emf}_2)}{r_1 + r_2}$ flows. See [\[link\]](#), for example, which shows a circuit exactly analogous to the battery charger discussed above. If two voltage sources in series with emfs in the same sense are connected to a load R_{load} , as in [\[link\]](#), then $I = \frac{(\text{emf}_1 + \text{emf}_2)}{r_1 + r_2 + R_{\text{load}}}$ flows.



These two voltage sources are connected in series with their emfs in opposition. Current flows in the direction of the greater emf and is limited

to $I = \frac{(\text{emf}_1 - \text{emf}_2)}{r_1 + r_2}$ by the sum of the internal resistances. (Note that each emf is represented by script E in the figure.)

A battery charger connected to a battery is an example of such a connection. The charger must have a larger emf than the battery to reverse current through it.



This schematic represents a flashlight with two cells (voltage sources) and a single bulb (load resistance) in series.

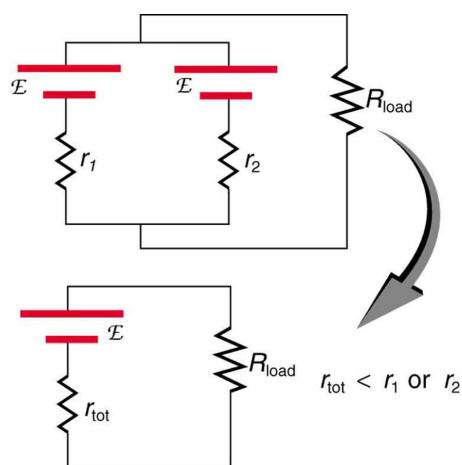
The current that flows is $I = \frac{(\text{emf}_1 + \text{emf}_2)}{r_1 + r_2 + R_{\text{load}}}$. (Note that each emf is represented by script E in the figure.)

Note:**Take-Home Experiment: Flashlight Batteries**

Find a flashlight that uses several batteries and find new and old batteries. Based on the discussions in this module, predict the brightness of the flashlight when different combinations of batteries are used. Do your predictions match what you observe? Now place new batteries in the flashlight and leave the flashlight switched on for several hours. Is the flashlight still quite bright? Do the same with the old batteries. Is the flashlight as bright when left on for the same length of time with old and new batteries? What does this say for the case when you are limited in the number of available new batteries?

[\[link\]](#) shows two voltage sources with identical emfs in parallel and connected to a load resistance. In this simple case, the total emf is the same as the individual emfs. But the total internal resistance is reduced, since the internal resistances are in parallel. The parallel connection thus can produce a larger current.

Here, $I = \frac{\text{emf}}{(r_{\text{tot}} + R_{\text{load}})}$ flows through the load, and r_{tot} is less than those of the individual batteries. For example, some diesel-powered cars use two 12-V batteries in parallel; they produce a total emf of 12 V but can deliver the larger current needed to start a diesel engine.



Two voltage sources
with identical emfs
(each labeled by script
E) connected in
parallel produce the
same emf but have a
smaller total internal
resistance than the
individual sources.
Parallel combinations
are often used to
deliver more current.
Here $I = \frac{\text{emf}}{(r_{\text{tot}} + R_{\text{load}})}$
flows through the
load.

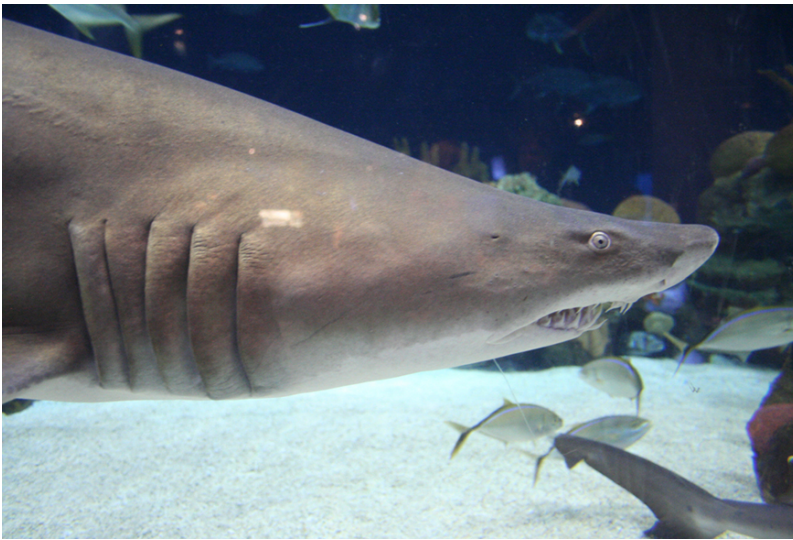
Animals as Electrical Detectors

A number of animals both produce and detect electrical signals. Fish, sharks, platypuses, and echidnas (spiny anteaters) all detect electric fields generated by nerve activity in prey. Electric eels produce their own emf through biological cells (electric organs) called electroplaques, which are arranged in both series and parallel as a set of batteries.

Electroplaques are flat, disk-like cells; those of the electric eel have a voltage of 0.15 V across each one. These cells are usually located toward the head or tail of the animal, although in the case of the electric eel, they are found along the entire body. The electroplaques in the South American eel are arranged in 140 rows, with each row stretching horizontally along the body and containing 5,000 electroplaques. This can yield an emf of approximately 600 V, and a current of 1 A—deadly.

The mechanism for detection of external electric fields is similar to that for producing nerve signals in the cell through depolarization and

repolarization—the movement of ions across the cell membrane. Within the fish, weak electric fields in the water produce a current in a gel-filled canal that runs from the skin to sensing cells, producing a nerve signal. The Australian platypus, one of the very few mammals that lay eggs, can detect fields of $30 \frac{\text{mV}}{\text{m}}$, while sharks have been found to be able to sense a field in their snouts as small as $100 \frac{\text{mV}}{\text{m}}$ ([link](#)). Electric eels use their own electric fields produced by the electroplaques to stun their prey or enemies.



Sand tiger sharks (*Carcharias taurus*), like this one at the Minnesota Zoo, use electroreceptors in their snouts to locate prey. (credit: Jim Winstead, Flickr)

Solar Cell Arrays

Another example dealing with multiple voltage sources is that of combinations of solar cells—wired in both series and parallel combinations to yield a desired voltage and current. Photovoltaic generation (PV), the conversion of sunlight directly into electricity, is based upon the

photoelectric effect, in which photons hitting the surface of a solar cell create an electric current in the cell.

Most solar cells are made from pure silicon—either as single-crystal silicon, or as a thin film of silicon deposited upon a glass or metal backing. Most single cells have a voltage output of about 0.5 V, while the current output is a function of the amount of sunlight upon the cell (the incident solar radiation—the insolation). Under bright noon sunlight, a current of about 100 mA/cm^2 of cell surface area is produced by typical single-crystal cells.

Individual solar cells are connected electrically in modules to meet electrical-energy needs. They can be wired together in series or in parallel—connected like the batteries discussed earlier. A solar-cell array or module usually consists of between 36 and 72 cells, with a power output of 50 W to 140 W.

The output of the solar cells is direct current. For most uses in a home, AC is required, so a device called an inverter must be used to convert the DC to AC. Any extra output can then be passed on to the outside electrical grid for sale to the utility.

Note:

Take-Home Experiment: Virtual Solar Cells

One can assemble a “virtual” solar cell array by using playing cards, or business or index cards, to represent a solar cell. Combinations of these cards in series and/or parallel can model the required array output. Assume each card has an output of 0.5 V and a current (under bright light) of 2 A. Using your cards, how would you arrange them to produce an output of 6 A at 3 V (18 W)?

Suppose you were told that you needed only 18 W (but no required voltage). Would you need more cards to make this arrangement?

Section Summary

- All voltage sources have two fundamental parts—a source of electrical energy that has a characteristic electromotive force (emf), and an internal resistance r .
- The emf is the potential difference of a source when no current is flowing.
- The numerical value of the emf depends on the source of potential difference.
- The internal resistance r of a voltage source affects the output voltage when a current flows.
- The voltage output of a device is called its terminal voltage V and is given by $V = \text{emf} - Ir$, where I is the electric current and is positive when flowing away from the positive terminal of the voltage source.
- When multiple voltage sources are in series, their internal resistances add and their emfs add algebraically.
- Solar cells can be wired in series or parallel to provide increased voltage or current, respectively.

Conceptual Questions

Exercise:

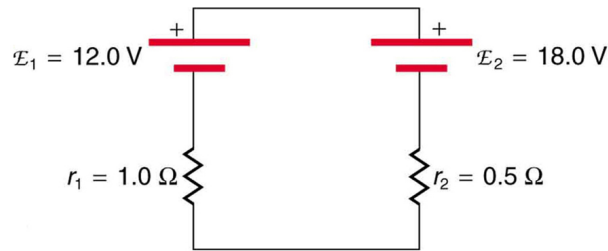
Problem:

Is every emf a potential difference? Is every potential difference an emf? Explain.

Exercise:

Problem:

Explain which battery is doing the charging and which is being charged in [\[link\]](#).



Exercise:

Problem:

Given a battery, an assortment of resistors, and a variety of voltage and current measuring devices, describe how you would determine the internal resistance of the battery.

Exercise:

Problem:

Two different 12-V automobile batteries on a store shelf are rated at 600 and 850 “cold cranking amps.” Which has the smallest internal resistance?

Exercise:

Problem:

What are the advantages and disadvantages of connecting batteries in series? In parallel?

Exercise:

Problem:

Semitractor trucks use four large 12-V batteries. The starter system requires 24 V, while normal operation of the truck’s other electrical components utilizes 12 V. How could the four batteries be connected to produce 24 V? To produce 12 V? Why is 24 V better than 12 V for starting the truck’s engine (a very heavy load)?

Problem Exercises

Exercise:**Problem:**

Standard automobile batteries have six lead-acid cells in series, creating a total emf of 12.0 V. What is the emf of an individual lead-acid cell?

Solution:

2.00 V

Exercise:**Problem:**

Carbon-zinc dry cells (sometimes referred to as non-alkaline cells) have an emf of 1.54 V, and they are produced as single cells or in various combinations to form other voltages. (a) How many 1.54-V cells are needed to make the common 9-V battery used in many small electronic devices? (b) What is the actual emf of the approximately 9-V battery? (c) Discuss how internal resistance in the series connection of cells will affect the terminal voltage of this approximately 9-V battery.

Exercise:**Problem:**

What is the output voltage of a 3.0000-V lithium cell in a digital wristwatch that draws 0.300 mA, if the cell's internal resistance is $2.00\ \Omega$?

Solution:

2.9994 V

Exercise:

Problem:

(a) What is the terminal voltage of a large 1.54-V carbon-zinc dry cell used in a physics lab to supply 2.00 A to a circuit, if the cell's internal resistance is $0.100\ \Omega$? (b) How much electrical power does the cell produce? (c) What power goes to its load?

Exercise:**Problem:**

What is the internal resistance of an automobile battery that has an emf of 12.0 V and a terminal voltage of 15.0 V while a current of 8.00 A is charging it?

Solution:

$0.375\ \Omega$

Exercise:**Problem:**

(a) Find the terminal voltage of a 12.0-V motorcycle battery having a $0.600\text{-}\Omega$ internal resistance, if it is being charged by a current of 10.0 A. (b) What is the output voltage of the battery charger?

Exercise:**Problem:**

A car battery with a 12-V emf and an internal resistance of $0.050\ \Omega$ is being charged with a current of 60 A. Note that in this process the battery is being charged. (a) What is the potential difference across its terminals? (b) At what rate is thermal energy being dissipated in the battery? (c) At what rate is electric energy being converted to chemical energy? (d) What are the answers to (a) and (b) when the battery is used to supply 60 A to the starter motor?

Exercise:

Problem:

The hot resistance of a flashlight bulb is $2.30\ \Omega$, and it is run by a 1.58-V alkaline cell having a $0.100\text{-}\Omega$ internal resistance. (a) What current flows? (b) Calculate the power supplied to the bulb using $I^2 R_{\text{bulb}}$. (c) Is this power the same as calculated using $\frac{V^2}{R_{\text{bulb}}}$?

Solution:

(a) 0.658 A

(b) 0.997 W

(c) 0.997 W; yes

Exercise:**Problem:**

The label on a portable radio recommends the use of rechargeable nickel-cadmium cells (nicads), although they have a 1.25-V emf while alkaline cells have a 1.58-V emf. The radio has a $3.20\text{-}\Omega$ resistance. (a) Draw a circuit diagram of the radio and its batteries. Now, calculate the power delivered to the radio. (b) When using Nicad cells each having an internal resistance of $0.0400\ \Omega$. (c) When using alkaline cells each having an internal resistance of $0.200\ \Omega$. (d) Does this difference seem significant, considering that the radio's effective resistance is lowered when its volume is turned up?

Exercise:

Problem:

An automobile starter motor has an equivalent resistance of $0.0500\ \Omega$ and is supplied by a 12.0-V battery with a $0.0100\text{-}\Omega$ internal resistance. (a) What is the current to the motor? (b) What voltage is applied to it? (c) What power is supplied to the motor? (d) Repeat these calculations for when the battery connections are corroded and add $0.0900\ \Omega$ to the circuit. (Significant problems are caused by even small amounts of unwanted resistance in low-voltage, high-current applications.)

Solution:

- (a) 200 A
- (b) 10.0 V
- (c) 2.00 kW
- (d) $0.1000\ \Omega$; 80.0 A, 4.0 V, 320 W

Exercise:**Problem:**

A child's electronic toy is supplied by three 1.58-V alkaline cells having internal resistances of $0.0200\ \Omega$ in series with a 1.53-V carbon-zinc dry cell having a $0.100\text{-}\Omega$ internal resistance. The load resistance is $10.0\ \Omega$. (a) Draw a circuit diagram of the toy and its batteries. (b) What current flows? (c) How much power is supplied to the load? (d) What is the internal resistance of the dry cell if it goes bad, resulting in only 0.500 W being supplied to the load?

Exercise:

Problem:

(a) What is the internal resistance of a voltage source if its terminal voltage drops by 2.00 V when the current supplied increases by 5.00 A? (b) Can the emf of the voltage source be found with the information supplied?

Solution:

(a) $0.400\ \Omega$

(b) No, there is only one independent equation, so only r can be found.

Exercise:**Problem:**

A person with body resistance between his hands of $10.0\ \text{k}\Omega$ accidentally grasps the terminals of a 20.0-kV power supply. (Do NOT do this!) (a) Draw a circuit diagram to represent the situation. (b) If the internal resistance of the power supply is $2000\ \Omega$, what is the current through his body? (c) What is the power dissipated in his body? (d) If the power supply is to be made safe by increasing its internal resistance, what should the internal resistance be for the maximum current in this situation to be 1.00 mA or less? (e) Will this modification compromise the effectiveness of the power supply for driving low-resistance devices? Explain your reasoning.

Exercise:**Problem:**

Electric fish generate current with biological cells called electroplaques, which are physiological emf devices. The electroplaques in the South American eel are arranged in 140 rows, each row stretching horizontally along the body and each containing 5000 electroplaques. Each electroplaque has an emf of 0.15 V and internal resistance of $0.25\ \Omega$. If the water surrounding the fish has resistance of $800\ \Omega$, how much current can the eel produce in water from near its head to near its tail?

Exercise:**Problem: Integrated Concepts**

A 12.0-V emf automobile battery has a terminal voltage of 16.0 V when being charged by a current of 10.0 A. (a) What is the battery's internal resistance? (b) What power is dissipated inside the battery? (c) At what rate (in $^{\circ}\text{C}/\text{min}$) will its temperature increase if its mass is 20.0 kg and it has a specific heat of $0.300 \text{ kcal}/\text{kg} \cdot ^{\circ}\text{C}$, assuming no heat escapes?

Exercise:**Problem: Unreasonable Results**

A 1.58-V alkaline cell with a $0.200\text{-}\Omega$ internal resistance is supplying 8.50 A to a load. (a) What is its terminal voltage? (b) What is the value of the load resistance? (c) What is unreasonable about these results? (d) Which assumptions are unreasonable or inconsistent?

Solution:

(a) -0.120 V

(b) $-1.41 \times 10^{-2} \Omega$

(c) Negative terminal voltage; negative load resistance.

(d) The assumption that such a cell could provide 8.50 A is inconsistent with its internal resistance.

Exercise:**Problem: Unreasonable Results**

(a) What is the internal resistance of a 1.54-V dry cell that supplies 1.00 W of power to a $15.0\text{-}\Omega$ bulb? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

electromotive force (emf)

the potential difference of a source of electricity when no current is flowing; measured in volts

internal resistance

the amount of resistance within the voltage source

potential difference

the difference in electric potential between two points in an electric circuit, measured in volts

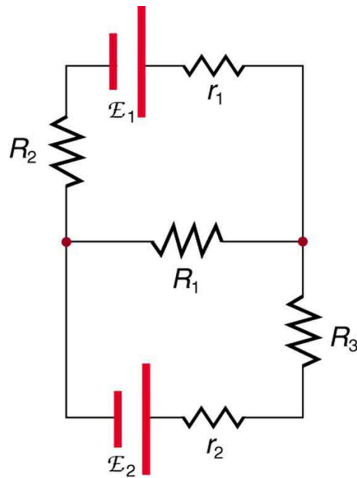
terminal voltage

the voltage measured across the terminals of a source of potential difference

Kirchhoff's Rules

- Analyze a complex circuit using Kirchhoff's rules, using the conventions for determining the correct signs of various terms.

Many complex circuits, such as the one in [\[link\]](#), cannot be analyzed with the series-parallel techniques developed in [Resistors in Series and Parallel](#) and [Electromotive Force: Terminal Voltage](#). There are, however, two circuit analysis rules that can be used to analyze any circuit, simple or complex. These rules are special cases of the laws of conservation of charge and conservation of energy. The rules are known as **Kirchhoff's rules**, after their inventor Gustav Kirchhoff (1824–1887).



This circuit cannot be reduced to a combination of series and parallel connections. Kirchhoff's rules, special applications of the laws of conservation of charge and energy, can be

used to analyze it. (Note: The script E in the figure represents electromotive force, emf.)

Note:

Kirchhoff's Rules

- Kirchhoff's first rule—the junction rule. The sum of all currents entering a junction must equal the sum of all currents leaving the junction.
- Kirchhoff's second rule—the loop rule. The algebraic sum of changes in potential around any closed circuit path (loop) must be zero.

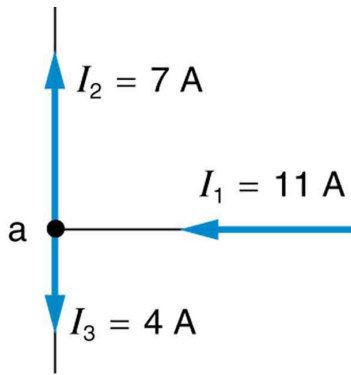
Explanations of the two rules will now be given, followed by problem-solving hints for applying Kirchhoff's rules, and a worked example that uses them.

Kirchhoff's First Rule

Kirchhoff's first rule (the **junction rule**) is an application of the conservation of charge to a junction; it is illustrated in [\[link\]](#). Current is the flow of charge, and charge is conserved; thus, whatever charge flows into the junction must flow out. Kirchhoff's first rule requires that $I_1 = I_2 + I_3$ (see figure). Equations like this can and will be used to analyze circuits and to solve circuit problems.

Note:**Making Connections: Conservation Laws**

Kirchhoff's rules for circuit analysis are applications of **conservation laws** to circuits. The first rule is the application of conservation of charge, while the second rule is the application of conservation of energy. Conservation laws, even used in a specific application, such as circuit analysis, are so basic as to form the foundation of that application.



$$I_1 = I_2 + I_3$$

The junction rule. The diagram shows an example of Kirchhoff's first rule where the sum of the currents into a junction equals the sum of the currents out of a junction. In this case, the current going into the junction splits and comes out as

two currents, so
that

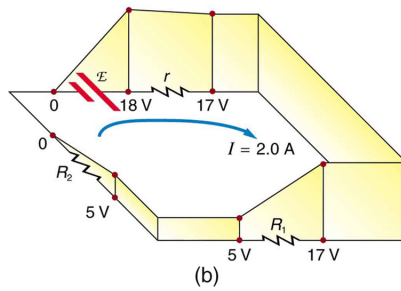
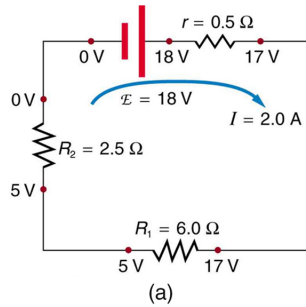
$$I_1 = I_2 + I_3.$$

Here I_1 must be
11 A, since I_2 is
7 A and I_3 is 4
A.

Kirchhoff's Second Rule

Kirchhoff's second rule (the **loop rule**) is an application of conservation of energy. The loop rule is stated in terms of potential, V , rather than potential energy, but the two are related since $PE_{\text{elec}} = qV$. Recall that **emf** is the potential difference of a source when no current is flowing. In a closed loop, whatever energy is supplied by emf must be transferred into other forms by devices in the loop, since there are no other ways in which energy can be transferred into or out of the circuit. [\[link\]](#) illustrates the changes in potential in a simple series circuit loop.

Kirchhoff's second rule requires $\text{emf} - Ir - IR_1 - IR_2 = 0$. Rearranged, this is $\text{emf} = Ir + IR_1 + IR_2$, which means the emf equals the sum of the IR (voltage) drops in the loop.



The loop rule. An example of Kirchhoff's second rule where the sum of the changes in potential around a closed loop must be zero. (a) In this standard schematic of a simple series circuit, the emf supplies 18 V, which is reduced to zero by the resistances, with 1 V across the internal resistance, and 12 V and 5 V across the two load resistances, for a total of 18 V. (b) This perspective view represents the

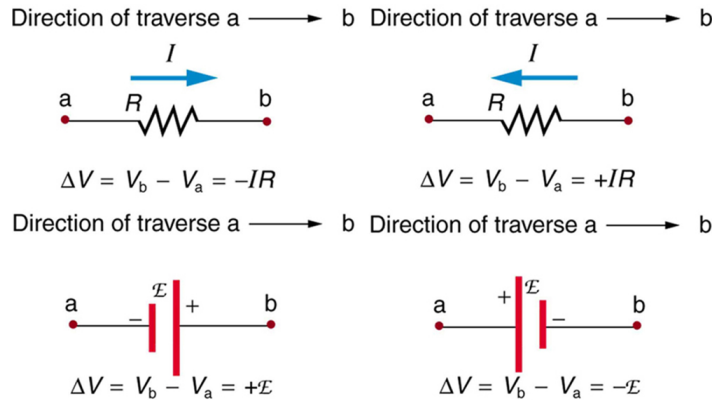
potential as something like a roller coaster, where charge is raised in potential by the emf and lowered by the resistances. (Note that the script E stands for emf.)

Applying Kirchhoff's Rules

By applying Kirchhoff's rules, we generate equations that allow us to find the unknowns in circuits. The unknowns may be currents, emfs, or resistances. Each time a rule is applied, an equation is produced. If there are as many independent equations as unknowns, then the problem can be solved. There are two decisions you must make when applying Kirchhoff's rules. These decisions determine the signs of various quantities in the equations you obtain from applying the rules.

1. When applying Kirchhoff's first rule, the junction rule, you must label the current in each branch and decide in what direction it is going. For example, in [\[link\]](#), [\[link\]](#), and [\[link\]](#), currents are labeled I_1 , I_2 , I_3 , and I , and arrows indicate their directions. There is no risk here, for if you choose the wrong direction, the current will be of the correct magnitude but negative.
2. When applying Kirchhoff's second rule, the loop rule, you must identify a closed loop and decide in which direction to go around it, clockwise or counterclockwise. For example, in [\[link\]](#) the loop was traversed in the same direction as the current (clockwise). Again, there is no risk; going around the circuit in the opposite direction reverses the sign of every term in the equation, which is like multiplying both sides of the equation by -1 .

[\[link\]](#) and the following points will help you get the plus or minus signs right when applying the loop rule. Note that the resistors and emfs are traversed by going from a to b. In many circuits, it will be necessary to construct more than one loop. In traversing each loop, one needs to be consistent for the sign of the change in potential. (See [\[link\]](#).)

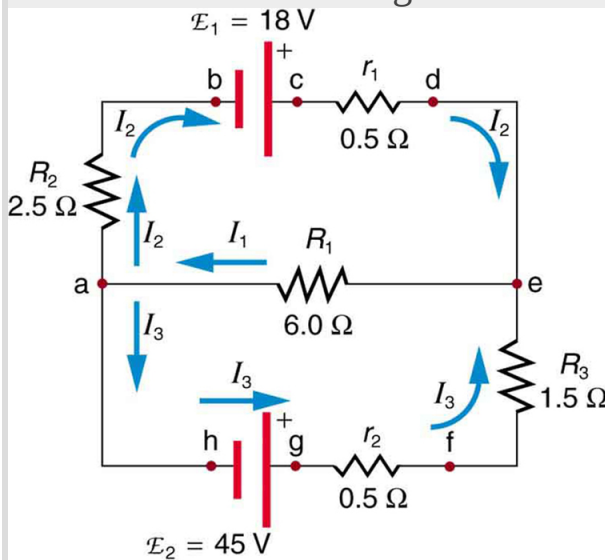


Each of these resistors and voltage sources is traversed from a to b. The potential changes are shown beneath each element and are explained in the text. (Note that the script E stands for emf.)

- When a resistor is traversed in the same direction as the current, the change in potential is $-IR$. (See [\[link\]](#).)
- When a resistor is traversed in the direction opposite to the current, the change in potential is $+IR$. (See [\[link\]](#).)
- When an emf is traversed from $-$ to $+$ (the same direction it moves positive charge), the change in potential is $+\text{emf}$. (See [\[link\]](#).)
- When an emf is traversed from $+$ to $-$ (opposite to the direction it moves positive charge), the change in potential is $-\text{emf}$. (See [\[link\]](#).)

Example:**Calculating Current: Using Kirchhoff's Rules**

Find the currents flowing in the circuit in [\[link\]](#).



This circuit is similar to that in [\[link\]](#), but the resistances and emfs are specified. (Each emf is denoted by script E.) The currents in each branch are labeled and assumed to move in the directions shown. This example uses Kirchhoff's rules to find the currents.

Strategy

This circuit is sufficiently complex that the currents cannot be found using Ohm's law and the series-parallel techniques—it is necessary to use Kirchhoff's rules. Currents have been labeled I_1 , I_2 , and I_3 in the figure and assumptions have been made about their directions. Locations on the diagram have been labeled with letters a through h. In the solution we will apply the junction and loop rules, seeking three independent equations to allow us to solve for the three unknown currents.

Solution

We begin by applying Kirchhoff's first or junction rule at point a. This gives

Equation:

$$I_1 = I_2 + I_3,$$

since I_1 flows into the junction, while I_2 and I_3 flow out. Applying the junction rule at e produces exactly the same equation, so that no new information is obtained. This is a single equation with three unknowns—three independent equations are needed, and so the loop rule must be applied.

Now we consider the loop abcdea. Going from a to b, we traverse R_2 in the same (assumed) direction of the current I_2 , and so the change in potential is $-I_2 R_2$. Then going from b to c, we go from $-$ to $+$, so that the change in potential is $+\text{emf}_1$. Traversing the internal resistance r_1 from c to d gives $-I_2 r_1$. Completing the loop by going from d to a again traverses a resistor in the same direction as its current, giving a change in potential of $-I_1 R_1$.

The loop rule states that the changes in potential sum to zero. Thus,

Equation:

$$-I_2 R_2 + \text{emf}_1 - I_2 r_1 - I_1 R_1 = -I_2(R_2 + r_1) + \text{emf}_1 - I_1 R_1 = 0.$$

Substituting values from the circuit diagram for the resistances and emf, and canceling the ampere unit gives

Equation:

$$-3I_2 + 18 - 6I_1 = 0.$$

Now applying the loop rule to aefgha (we could have chosen abcdefgha as well) similarly gives

Equation:

$$+ I_1 R_1 + I_3 R_3 + I_3 r_2 - \text{emf}_2 = +I_1 R_1 + I_3(R_3 + r_2) - \text{emf}_2 = 0.$$

Note that the signs are reversed compared with the other loop, because elements are traversed in the opposite direction. With values entered, this becomes

Equation:

$$+ 6I_1 + 2I_3 - 45 = 0.$$

These three equations are sufficient to solve for the three unknown currents. First, solve the second equation for I_2 :

Equation:

$$I_2 = 6 - 2I_1.$$

Now solve the third equation for I_3 :

Equation:

$$I_3 = 22.5 - 3I_1.$$

Substituting these two new equations into the first one allows us to find a value for I_1 :

Equation:

$$I_1 = I_2 + I_3 = (6 - 2I_1) + (22.5 - 3I_1) = 28.5 - 5I_1.$$

Combining terms gives

Equation:

$$6I_1 = 28.5, \text{ and}$$

Equation:

$$I_1 = 4.75 \text{ A.}$$

Substituting this value for I_1 back into the fourth equation gives

Equation:

$$I_2 = 6 - 2I_1 = 6 - 9.50$$

Equation:

$$I_2 = -3.50 \text{ A.}$$

The minus sign means I_2 flows in the direction opposite to that assumed in [\[link\]](#).

Finally, substituting the value for I_1 into the fifth equation gives

Equation:

$$I_3 = 22.5 - 3I_1 = 22.5 - 14.25$$

Equation:

$$I_3 = 8.25 \text{ A.}$$

Discussion

Just as a check, we note that indeed $I_1 = I_2 + I_3$. The results could also have been checked by entering all of the values into the equation for the abcdefgha loop.

Note:**Problem-Solving Strategies for Kirchhoff's Rules**

1. Make certain there is a clear circuit diagram on which you can label all known and unknown resistances, emfs, and currents. If a current is unknown, you must assign it a direction. This is necessary for determining the signs of potential changes. If you assign the direction incorrectly, the current will be found to have a negative value—no harm done.
2. Apply the junction rule to any junction in the circuit. Each time the junction rule is applied, you should get an equation with a current that does not appear in a previous application—if not, then the equation is redundant.
3. Apply the loop rule to as many loops as needed to solve for the unknowns in the problem. (There must be as many independent equations as unknowns.) To apply the loop rule, you must choose a direction to go around the loop. Then carefully and consistently determine the signs of the potential changes for each element using the four bulleted points discussed above in conjunction with [\[link\]](#).
4. Solve the simultaneous equations for the unknowns. This may involve many algebraic steps, requiring careful checking and rechecking.
5. Check to see whether the answers are reasonable and consistent. The numbers should be of the correct order of magnitude, neither

exceedingly large nor vanishingly small. The signs should be reasonable—for example, no resistance should be negative. Check to see that the values obtained satisfy the various equations obtained from applying the rules. The currents should satisfy the junction rule, for example.

The material in this section is correct in theory. We should be able to verify it by making measurements of current and voltage. In fact, some of the devices used to make such measurements are straightforward applications of the principles covered so far and are explored in the next modules. As we shall see, a very basic, even profound, fact results—making a measurement alters the quantity being measured.

Exercise:

Check Your Understanding

Problem:

Can Kirchhoff's rules be applied to simple series and parallel circuits or are they restricted for use in more complicated circuits that are not combinations of series and parallel?

Solution:

Kirchhoff's rules can be applied to any circuit since they are applications to circuits of two conservation laws. Conservation laws are the most broadly applicable principles in physics. It is usually mathematically simpler to use the rules for series and parallel in simpler circuits so we emphasize Kirchhoff's rules for use in more complicated situations. But the rules for series and parallel can be derived from Kirchhoff's rules. Moreover, Kirchhoff's rules can be expanded to devices other than resistors and emfs, such as capacitors, and are one of the basic analysis devices in circuit analysis.

Section Summary

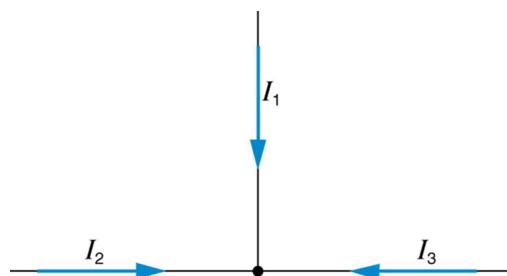
- Kirchhoff's rules can be used to analyze any circuit, simple or complex.
- Kirchhoff's first rule—the junction rule: The sum of all currents entering a junction must equal the sum of all currents leaving the junction.
- Kirchhoff's second rule—the loop rule: The algebraic sum of changes in potential around any closed circuit path (loop) must be zero.
- The two rules are based, respectively, on the laws of conservation of charge and energy.
- When calculating potential and current using Kirchhoff's rules, a set of conventions must be followed for determining the correct signs of various terms.
- The simpler series and parallel rules are special cases of Kirchhoff's rules.

Conceptual Questions

Exercise:

Problem:

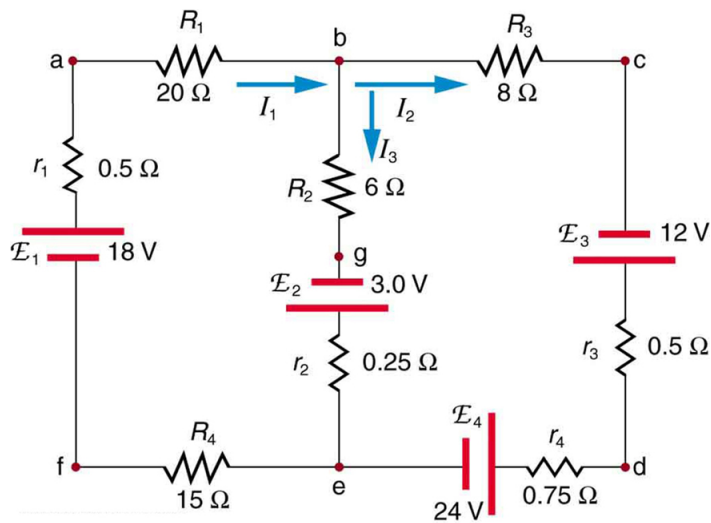
Can all of the currents going into the junction in [\[link\]](#) be positive? Explain.



Exercise:

Problem:

Apply the junction rule to junction b in [\[link\]](#). Is any new information gained by applying the junction rule at e? (In the figure, each emf is represented by script E.)

**Exercise:****Problem:**

(a) What is the potential difference going from point a to point b in [\[link\]](#)? (b) What is the potential difference going from c to b? (c) From e to g? (d) From e to d?

Exercise:

Problem: Apply the loop rule to loop afedcba in [\[link\]](#).

Exercise:

Problem: Apply the loop rule to loops abgefa and cbgedc in [\[link\]](#).

Problem Exercises

Exercise:

Problem: Apply the loop rule to loop abcdefgha in [\[link\]](#).

Solution:

Equation:

$$-I_2 R_2 + \text{emf}_1 - I_2 r_1 + I_3 R_3 + I_3 r_2 - \text{emf}_2 = 0$$

Exercise:

Problem: Apply the loop rule to loop aedcba in [\[link\]](#).

Exercise:

Problem:

Verify the second equation in [\[link\]](#) by substituting the values found for the currents I_1 and I_2 .

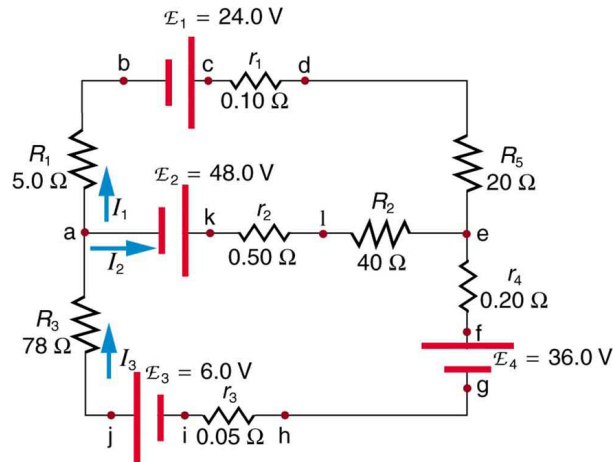
Exercise:

Problem:

Verify the third equation in [\[link\]](#) by substituting the values found for the currents I_1 and I_3 .

Exercise:

Problem: Apply the junction rule at point a in [\[link\]](#).



Solution:

Equation:

$$I_3 = I_1 + I_2$$

Exercise:

Problem: Apply the loop rule to loop abcdefghija in [\[link\]](#).

Exercise:

Problem: Apply the loop rule to loop akledcba in [\[link\]](#).

Solution:

Equation:

$$\text{emf}_2 - I_2 r_2 - I_2 R_2 + I_1 R_5 + I_1 r_1 - \text{emf}_1 + I_1 R_1 = 0$$

Exercise:

Problem:

Find the currents flowing in the circuit in [\[link\]](#). Explicitly show how you follow the steps in the [Problem-Solving Strategies for Series and Parallel Resistors](#).

Exercise:

Problem:

Solve [\[link\]](#), but use loop abcdefgha instead of loop akledcba. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Series and Parallel Resistors](#).

Solution:

(a) $I_1 = 4.75 \text{ A}$

(b) $I_2 = -3.5 \text{ A}$

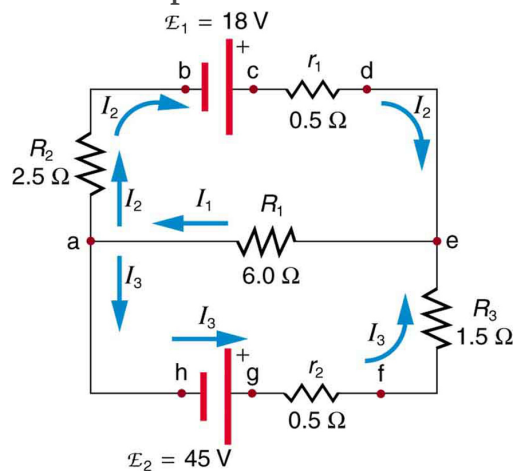
(c) $I_3 = 8.25 \text{ A}$

Exercise:

Problem: Find the currents flowing in the circuit in [\[link\]](#).

Exercise:**Problem: Unreasonable Results**

Consider the circuit in [\[link\]](#), and suppose that the emfs are unknown and the currents are given to be $I_1 = 5.00 \text{ A}$, $I_2 = 3.0 \text{ A}$, and $I_3 = -2.00 \text{ A}$. (a) Could you find the emfs? (b) What is wrong with the assumptions?



Solution:

- (a) No, you would get inconsistent equations to solve.
- (b) $I_1 \neq I_2 + I_3$. The assumed currents violate the junction rule.

Glossary**Kirchhoff's rules**

a set of two rules, based on conservation of charge and energy, governing current and changes in potential in an electric circuit

junction rule

Kirchhoff's first rule, which applies the conservation of charge to a junction; current is the flow of charge; thus, whatever charge flows into the junction must flow out; the rule can be stated $I_1 = I_2 + I_3$

loop rule

Kirchhoff's second rule, which states that in a closed loop, whatever energy is supplied by emf must be transferred into other forms by devices in the loop, since there are no other ways in which energy can be transferred into or out of the circuit. Thus, the emf equals the sum of the IR (voltage) drops in the loop and can be stated:
$$\text{emf} = Ir + IR_1 + IR_2$$

conservation laws

require that energy and charge be conserved in a system

DC Voltmeters and Ammeters

- Explain why a voltmeter must be connected in parallel with the circuit.
- Draw a diagram showing an ammeter correctly connected in a circuit.
- Describe how a galvanometer can be used as either a voltmeter or an ammeter.
- Find the resistance that must be placed in series with a galvanometer to allow it to be used as a voltmeter with a given reading.
- Explain why measuring the voltage or current in a circuit can never be exact.

Voltmeters measure voltage, whereas **ammeters** measure current. Some of the meters in automobile dashboards, digital cameras, cell phones, and tuner-amplifiers are voltmeters or ammeters. (See [\[link\]](#).) The internal construction of the simplest of these meters and how they are connected to the system they monitor give further insight into applications of series and parallel connections.

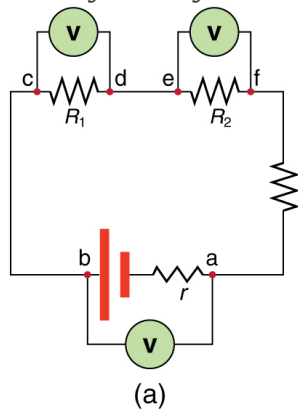


The fuel and temperature gauges (far right and far left, respectively) in this 1996 Volkswagen are voltmeters that register the voltage output of “sender” units, which are hopefully proportional to the amount of gasoline in the tank and the engine

temperature. (credit:
Christian Giersing)

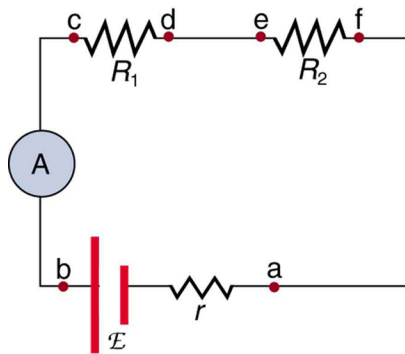
Voltmeters are connected in parallel with whatever device's voltage is to be measured. A parallel connection is used because objects in parallel experience the same potential difference. (See [\[link\]](#), where the voltmeter is represented by the symbol V.)

Ammeters are connected in series with whatever device's current is to be measured. A series connection is used because objects in series have the same current passing through them. (See [\[link\]](#), where the ammeter is represented by the symbol A.)



(a) To measure potential differences in this series circuit, the voltmeter (V) is placed in parallel with the

voltage source or either of the resistors. Note that terminal voltage is measured between points a and b. It is not possible to connect the voltmeter directly across the emf without including its internal resistance, r . (b) A digital voltmeter in use.
(credit: Messtechniker, Wikimedia Commons)



An ammeter (A) is placed in series to measure current. All of the current in this circuit flows through the meter.

The ammeter would have the same reading if located between points d and e or between points f and a as it does in the position shown. (Note that the script

capital E stands for
emf, and r stands
for the internal
resistance of the
source of potential
difference.)

Analog Meters: Galvanometers

Analog meters have a needle that swivels to point at numbers on a scale, as opposed to **digital meters**, which have numerical readouts similar to a hand-held calculator. The heart of most analog meters is a device called a **galvanometer**, denoted by G . Current flow through a galvanometer, I_G , produces a proportional needle deflection. (This deflection is due to the force of a magnetic field upon a current-carrying wire.)

The two crucial characteristics of a given galvanometer are its resistance and current sensitivity. **Current sensitivity** is the current that gives a **full-scale deflection** of the galvanometer's needle, the maximum current that the instrument can measure. For example, a galvanometer with a current sensitivity of $50\ \mu\text{A}$ has a maximum deflection of its needle when $50\ \mu\text{A}$ flows through it, reads half-scale when $25\ \mu\text{A}$ flows through it, and so on.

If such a galvanometer has a $25\text{-}\Omega$ resistance, then a voltage of only $V = IR = (50\ \mu\text{A})(25\ \Omega) = 1.25\ \text{mV}$ produces a full-scale reading. By connecting resistors to this galvanometer in different ways, you can use it as either a voltmeter or ammeter that can measure a broad range of voltages or currents.

Galvanometer as Voltmeter

[\[link\]](#) shows how a galvanometer can be used as a voltmeter by connecting it in series with a large resistance, R . The value of the resistance R is

determined by the maximum voltage to be measured. Suppose you want 10 V to produce a full-scale deflection of a voltmeter containing a $25\text{-}\Omega$ galvanometer with a $50\text{-}\mu\text{A}$ sensitivity. Then 10 V applied to the meter must produce a current of $50\text{ }\mu\text{A}$. The total resistance must be

Equation:

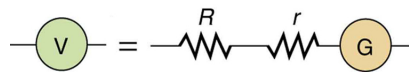
$$R_{\text{tot}} = R + r = \frac{V}{I} = \frac{10\text{ V}}{50\text{ }\mu\text{A}} = 200\text{ k}\Omega, \text{ or}$$

Equation:

$$R = R_{\text{tot}} - r = 200\text{ k}\Omega - 25\text{ }\Omega \approx 200\text{ k}\Omega.$$

(R is so large that the galvanometer resistance, r , is nearly negligible.) Note that 5 V applied to this voltmeter produces a half-scale deflection by producing a $25\text{-}\mu\text{A}$ current through the meter, and so the voltmeter's reading is proportional to voltage as desired.

This voltmeter would not be useful for voltages less than about half a volt, because the meter deflection would be small and difficult to read accurately. For other voltage ranges, other resistances are placed in series with the galvanometer. Many meters have a choice of scales. That choice involves switching an appropriate resistance into series with the galvanometer.



A large resistance
 R placed in series
 with a
 galvanometer G
 produces a
 voltmeter, the full-
 scale deflection of
 which depends on
 the choice of R .

The larger the voltage to be measured, the larger R must be. (Note that r represents the internal resistance of the galvanometer.)

Galvanometer as Ammeter

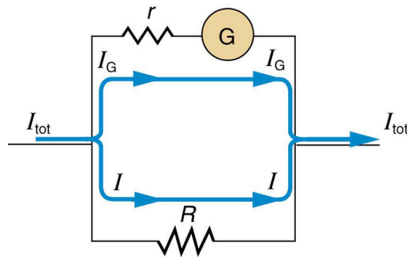
The same galvanometer can also be made into an ammeter by placing it in parallel with a small resistance R , often called the **shunt resistance**, as shown in [\[link\]](#). Since the shunt resistance is small, most of the current passes through it, allowing an ammeter to measure currents much greater than those producing a full-scale deflection of the galvanometer.

Suppose, for example, an ammeter is needed that gives a full-scale deflection for 1.0 A, and contains the same $25\text{-}\Omega$ galvanometer with its $50\text{-}\mu\text{A}$ sensitivity. Since R and r are in parallel, the voltage across them is the same.

These IR drops are $IR = I_G r$ so that $IR = \frac{I_G}{I} = \frac{R}{r}$. Solving for R , and noting that I_G is $50\text{ }\mu\text{A}$ and I is 0.999950 A , we have

Equation:

$$R = r \frac{I_G}{I} = (25\text{ }\Omega) \frac{50\text{ }\mu\text{A}}{0.999950\text{ A}} = 1.25 \times 10^{-3}\text{ }\Omega.$$



A small shunt resistance R placed in parallel with a galvanometer G produces an ammeter, the full-scale deflection of which depends on the choice of R .

The larger the current to be measured, the smaller R must be. Most of the current (I) flowing through the meter is shunted through R to protect the galvanometer.

(Note that r represents the internal resistance of the galvanometer.)

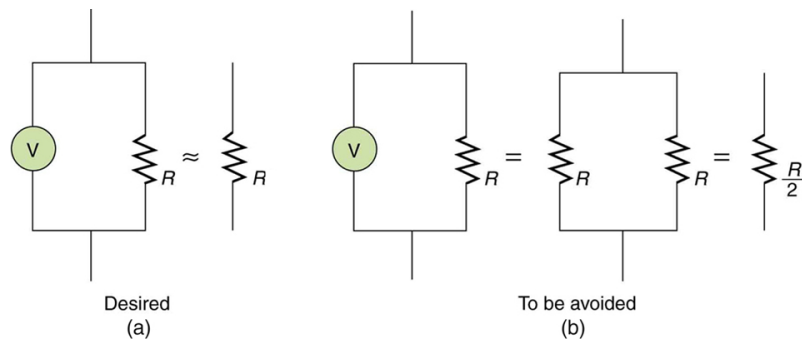
Ammeters may also have multiple scales for greater flexibility in application. The various scales are

achieved by
switching various
shunt resistances in
parallel with the
galvanometer—the
greater the
maximum current
to be measured, the
smaller the shunt
resistance must be.

Taking Measurements Alters the Circuit

When you use a voltmeter or ammeter, you are connecting another resistor to an existing circuit and, thus, altering the circuit. Ideally, voltmeters and ammeters do not appreciably affect the circuit, but it is instructive to examine the circumstances under which they do or do not interfere.

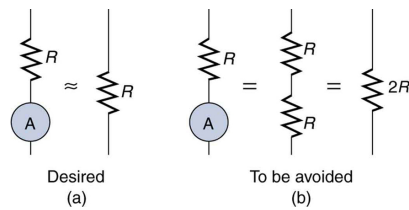
First, consider the voltmeter, which is always placed in parallel with the device being measured. Very little current flows through the voltmeter if its resistance is a few orders of magnitude greater than the device, and so the circuit is not appreciably affected. (See [\[link\]](#)(a).) (A large resistance in parallel with a small one has a combined resistance essentially equal to the small one.) If, however, the voltmeter's resistance is comparable to that of the device being measured, then the two in parallel have a smaller resistance, appreciably affecting the circuit. (See [\[link\]](#)(b).) The voltage across the device is not the same as when the voltmeter is out of the circuit.



(a) A voltmeter having a resistance much larger than the device ($R_{\text{Voltmeter}} \gg R$) with which it is in parallel produces a parallel resistance essentially the same as the device and does not appreciably affect the circuit being measured. (b) Here the voltmeter has the same resistance as the device ($R_{\text{Voltmeter}} \cong R$), so that the parallel resistance is half of what it is when the voltmeter is not connected. This is an example of a significant alteration of the circuit and is to be avoided.

An ammeter is placed in series in the branch of the circuit being measured, so that its resistance adds to that branch. Normally, the ammeter's resistance is very small compared with the resistances of the devices in the circuit, and so the extra resistance is negligible. (See [\[link\]](#)(a).) However, if very small load resistances are involved, or if the ammeter is not as low in resistance as it should be, then the total series resistance is significantly greater, and the current in the branch being measured is reduced. (See [\[link\]](#)(b).)

A practical problem can occur if the ammeter is connected incorrectly. If it was put in parallel with the resistor to measure the current in it, you could possibly damage the meter; the low resistance of the ammeter would allow most of the current in the circuit to go through the galvanometer, and this current would be larger since the effective resistance is smaller.



(a) An ammeter normally has such a small resistance that the total series resistance in the branch being measured is not appreciably increased. The circuit is essentially unaltered compared with when the ammeter is absent.

(b) Here the ammeter's resistance is the same as that of the branch, so that the total resistance is doubled and the current is half what it is without the ammeter. This significant alteration of the circuit is to be avoided.

One solution to the problem of voltmeters and ammeters interfering with the circuits being measured is to use galvanometers with greater sensitivity.

This allows construction of voltmeters with greater resistance and ammeters with smaller resistance than when less sensitive galvanometers are used.

There are practical limits to galvanometer sensitivity, but it is possible to get analog meters that make measurements accurate to a few percent. Note that the inaccuracy comes from altering the circuit, not from a fault in the meter.

Note:**Connections: Limits to Knowledge**

Making a measurement alters the system being measured in a manner that produces uncertainty in the measurement. For macroscopic systems, such as the circuits discussed in this module, the alteration can usually be made negligibly small, but it cannot be eliminated entirely. For submicroscopic systems, such as atoms, nuclei, and smaller particles, measurement alters the system in a manner that cannot be made arbitrarily small. This actually limits knowledge of the system—even limiting what nature can know about itself. We shall see profound implications of this when the Heisenberg uncertainty principle is discussed in the modules on quantum mechanics.

There is another measurement technique based on drawing no current at all and, hence, not altering the circuit at all. These are called null measurements and are the topic of [Null Measurements](#). Digital meters that employ solid-state electronics and null measurements can attain accuracies of one part in 10^6 .

Exercise:**Check Your Understanding****Problem:**

Digital meters are able to detect smaller currents than analog meters employing galvanometers. How does this explain their ability to measure voltage and current more accurately than analog meters?

Solution:

Since digital meters require less current than analog meters, they alter the circuit less than analog meters. Their resistance as a voltmeter can be far greater than an analog meter, and their resistance as an ammeter can be far less than an analog meter. Consult [\[link\]](#) and [\[link\]](#) and their discussion in the text.

Note:

PhET Explorations: Circuit Construction Kit (DC Only), Virtual Lab
Stimulate a neuron and monitor what happens. Pause, rewind, and move forward in time in order to observe the ions as they move across the neuron membrane.

[Circuit
Construction
Kit \(DC
Only\),
Virtual Lab](#)

Section Summary

- Voltmeters measure voltage, and ammeters measure current.
- A voltmeter is placed in parallel with the voltage source to receive full voltage and must have a large resistance to limit its effect on the circuit.
- An ammeter is placed in series to get the full current flowing through a branch and must have a small resistance to limit its effect on the circuit.

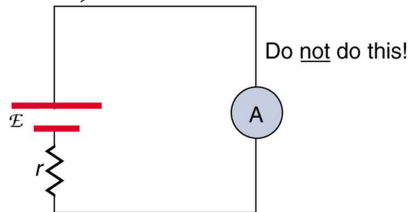
- Both can be based on the combination of a resistor and a galvanometer, a device that gives an analog reading of current.
- Standard voltmeters and ammeters alter the circuit being measured and are thus limited in accuracy.

Conceptual Questions

Exercise:

Problem:

Why should you not connect an ammeter directly across a voltage source as shown in [\[link\]](#)? (Note that script E in the figure stands for emf.)



Exercise:

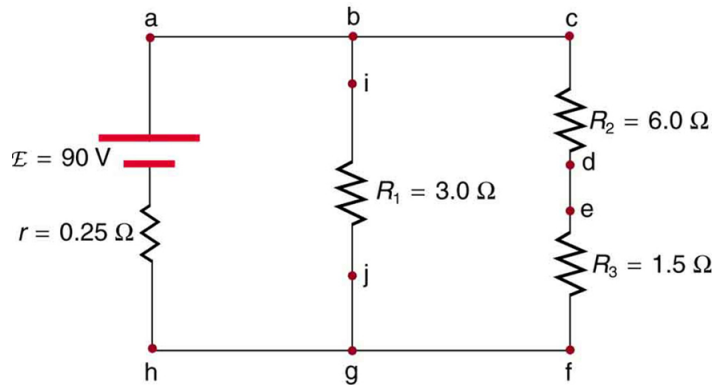
Problem:

Suppose you are using a multimeter (one designed to measure a range of voltages, currents, and resistances) to measure current in a circuit and you inadvertently leave it in a voltmeter mode. What effect will the meter have on the circuit? What would happen if you were measuring voltage but accidentally put the meter in the ammeter mode?

Exercise:

Problem:

Specify the points to which you could connect a voltmeter to measure the following potential differences in [\[link\]](#): (a) the potential difference of the voltage source; (b) the potential difference across R_1 ; (c) across R_2 ; (d) across R_3 ; (e) across R_2 and R_3 . Note that there may be more than one answer to each part.



Exercise:

Problem:

To measure currents in [\[link\]](#), you would replace a wire between two points with an ammeter. Specify the points between which you would place an ammeter to measure the following: (a) the total current; (b) the current flowing through R_1 ; (c) through R_2 ; (d) through R_3 . Note that there may be more than one answer to each part.

Problem Exercises

Exercise:

Problem:

What is the sensitivity of the galvanometer (that is, what current gives a full-scale deflection) inside a voltmeter that has a $1.00\text{-M}\Omega$ resistance on its 30.0-V scale?

Solution:

$$30\ \mu\text{A}$$

Exercise:

Problem:

What is the sensitivity of the galvanometer (that is, what current gives a full-scale deflection) inside a voltmeter that has a $25.0\text{-k}\Omega$ resistance on its 100-V scale?

Exercise:**Problem:**

Find the resistance that must be placed in series with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as a voltmeter with a 0.100-V full-scale reading.

Solution:

$1.98\text{ k}\Omega$

Exercise:**Problem:**

Find the resistance that must be placed in series with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as a voltmeter with a 3000-V full-scale reading. Include a circuit diagram with your solution.

Exercise:**Problem:**

Find the resistance that must be placed in parallel with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as an ammeter with a 10.0-A full-scale reading. Include a circuit diagram with your solution.

Solution:**Equation:**

$$1.25 \times 10^{-4} \Omega$$

Exercise:**Problem:**

Find the resistance that must be placed in parallel with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as an ammeter with a 300-mA full-scale reading.

Exercise:**Problem:**

Find the resistance that must be placed in series with a $10.0\text{-}\Omega$ galvanometer having a $100\text{-}\mu\text{A}$ sensitivity to allow it to be used as a voltmeter with: (a) a 300-V full-scale reading, and (b) a 0.300-V full-scale reading.

Solution:

(a) $3.00\text{ M}\Omega$

(b) $2.99\text{ k}\Omega$

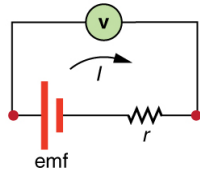
Exercise:**Problem:**

Find the resistance that must be placed in parallel with a $10.0\text{-}\Omega$ galvanometer having a $100\text{-}\mu\text{A}$ sensitivity to allow it to be used as an ammeter with: (a) a 20.0-A full-scale reading, and (b) a 100-mA full-scale reading.

Exercise:

Problem:

Suppose you measure the terminal voltage of a 1.585-V alkaline cell having an internal resistance of $0.100\ \Omega$ by placing a $1.00\text{-k}\Omega$ voltmeter across its terminals. (See [\[link\]](#).) (a) What current flows? (b) Find the terminal voltage. (c) To see how close the measured terminal voltage is to the emf, calculate their ratio.



Solution:

(a) 1.58 mA

(b) 1.5848 V (need four digits to see the difference)

(c) 0.99990 (need five digits to see the difference from unity)

Exercise:**Problem:**

Suppose you measure the terminal voltage of a 3.200-V lithium cell having an internal resistance of $5.00\ \Omega$ by placing a $1.00\text{-k}\Omega$ voltmeter across its terminals. (a) What current flows? (b) Find the terminal voltage. (c) To see how close the measured terminal voltage is to the emf, calculate their ratio.

Exercise:**Problem:**

A certain ammeter has a resistance of $5.00 \times 10^{-5}\ \Omega$ on its 3.00-A scale and contains a $10.0\text{-}\Omega$ galvanometer. What is the sensitivity of the galvanometer?

Solution:

15.0 μA

Exercise:

Problem:

A 1.00-M Ω voltmeter is placed in parallel with a 75.0-k Ω resistor in a circuit. (a) Draw a circuit diagram of the connection. (b) What is the resistance of the combination? (c) If the voltage across the combination is kept the same as it was across the 75.0-k Ω resistor alone, what is the percent increase in current? (d) If the current through the combination is kept the same as it was through the 75.0-k Ω resistor alone, what is the percentage decrease in voltage? (e) Are the changes found in parts (c) and (d) significant? Discuss.

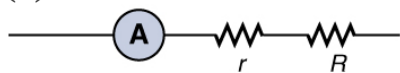
Exercise:

Problem:

A 0.0200- Ω ammeter is placed in series with a 10.00- Ω resistor in a circuit. (a) Draw a circuit diagram of the connection. (b) Calculate the resistance of the combination. (c) If the voltage is kept the same across the combination as it was through the 10.00- Ω resistor alone, what is the percent decrease in current? (d) If the current is kept the same through the combination as it was through the 10.00- Ω resistor alone, what is the percent increase in voltage? (e) Are the changes found in parts (c) and (d) significant? Discuss.

Solution:

(a)



(b) 10.02 Ω

(c) 0.9980, or a 2.0×10^{-1} percent decrease

(d) 1.002, or a 2.0×10^{-1} percent increase

(e) Not significant.

Exercise:

Problem: Unreasonable Results

Suppose you have a $40.0\text{-}\Omega$ galvanometer with a $25.0\text{-}\mu\text{A}$ sensitivity.

(a) What resistance would you put in series with it to allow it to be used as a voltmeter that has a full-scale deflection for 0.500 mV ? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

(a) What resistance would you put in parallel with a $40.0\text{-}\Omega$ galvanometer having a $25.0\text{-}\mu\text{A}$ sensitivity to allow it to be used as an ammeter that has a full-scale deflection for $10.0\text{-}\mu\text{A}$? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) $-66.7\text{ }\Omega$

(b) You can't have negative resistance.

(c) It is unreasonable that I_G is greater than I_{tot} (see [\[link\]](#)). You cannot achieve a full-scale deflection using a current less than the sensitivity of the galvanometer.

Glossary

voltmeter

an instrument that measures voltage

ammeter

an instrument that measures current

analog meter

a measuring instrument that gives a readout in the form of a needle movement over a marked gauge

digital meter

a measuring instrument that gives a readout in a digital form

galvanometer

an analog measuring device, denoted by G , that measures current flow using a needle deflection caused by a magnetic field force acting upon a current-carrying wire

current sensitivity

the maximum current that a galvanometer can read

full-scale deflection

the maximum deflection of a galvanometer needle, also known as current sensitivity; a galvanometer with a full-scale deflection of $50\ \mu\text{A}$ has a maximum deflection of its needle when $50\ \mu\text{A}$ flows through it

shunt resistance

a small resistance R placed in parallel with a galvanometer G to produce an ammeter; the larger the current to be measured, the smaller R must be; most of the current flowing through the meter is shunted through R to protect the galvanometer

Null Measurements

- Explain why a null measurement device is more accurate than a standard voltmeter or ammeter.
- Demonstrate how a Wheatstone bridge can be used to accurately calculate the resistance in a circuit.

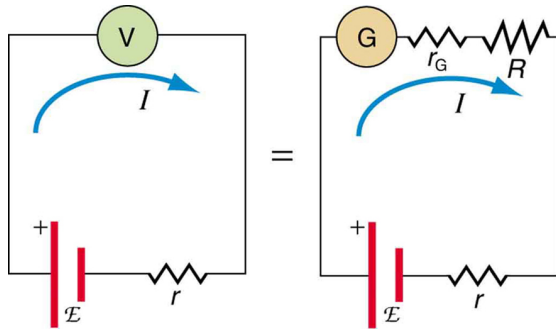
Standard measurements of voltage and current alter the circuit being measured, introducing uncertainties in the measurements. Voltmeters draw some extra current, whereas ammeters reduce current flow. **Null measurements** balance voltages so that there is no current flowing through the measuring device and, therefore, no alteration of the circuit being measured.

Null measurements are generally more accurate but are also more complex than the use of standard voltmeters and ammeters, and they still have limits to their precision. In this module, we shall consider a few specific types of null measurements, because they are common and interesting, and they further illuminate principles of electric circuits.

The Potentiometer

Suppose you wish to measure the emf of a battery. Consider what happens if you connect the battery directly to a standard voltmeter as shown in [\[link\]](#). (Once we note the problems with this measurement, we will examine a null measurement that improves accuracy.) As discussed before, the actual quantity measured is the terminal voltage V , which is related to the emf of the battery by $V = \text{emf} - Ir$, where I is the current that flows and r is the internal resistance of the battery.

The emf could be accurately calculated if r were very accurately known, but it is usually not. If the current I could be made zero, then $V = \text{emf}$, and so emf could be directly measured. However, standard voltmeters need a current to operate; thus, another technique is needed.



An analog voltmeter attached to a battery draws a small but nonzero current and measures a terminal voltage that differs from the emf of the battery. (Note that the script capital \mathcal{E} symbolizes electromotive force, or emf.) Since the internal resistance of the battery is not known precisely, it is not possible to calculate the emf precisely.

A **potentiometer** is a null measurement device for measuring potentials (voltages). (See [\[link\]](#).) A voltage source is connected to a resistor R , say, a long wire, and passes a constant current through it. There is a steady drop in potential (an IR drop) along the wire, so that a variable potential can be obtained by making contact at varying locations along the wire.

[\[link\]](#)(b) shows an unknown emf_x (represented by script \mathcal{E}_x in the figure) connected in series with a galvanometer. Note that emf_x opposes the other voltage source. The location of the contact point (see the arrow on the drawing) is adjusted until the galvanometer reads zero. When the galvanometer reads zero, $\text{emf}_x = IR_x$, where R_x is the resistance of the section of wire up to the contact point. Since no current flows through the

galvanometer, none flows through the unknown emf, and so emf_x is directly sensed.

Now, a very precisely known standard emf_s is substituted for emf_x , and the contact point is adjusted until the galvanometer again reads zero, so that $\text{emf}_s = IR_s$. In both cases, no current passes through the galvanometer, and so the current I through the long wire is the same. Upon taking the ratio $\frac{\text{emf}_x}{\text{emf}_s}$, I cancels, giving

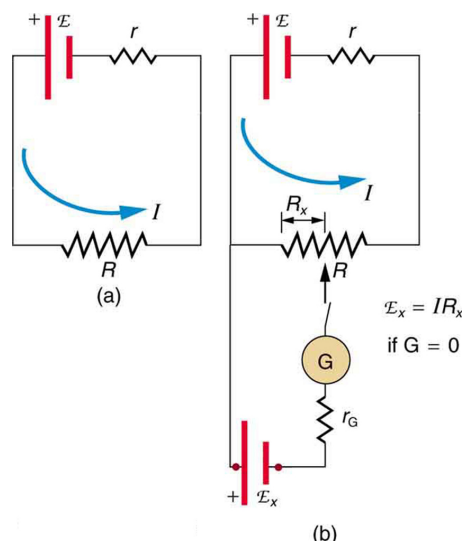
Equation:

$$\frac{\text{emf}_x}{\text{emf}_s} = \frac{IR_x}{IR_s} = \frac{R_x}{R_s}.$$

Solving for emf_x gives

Equation:

$$\text{emf}_x = \text{emf}_s \frac{R_x}{R_s}.$$



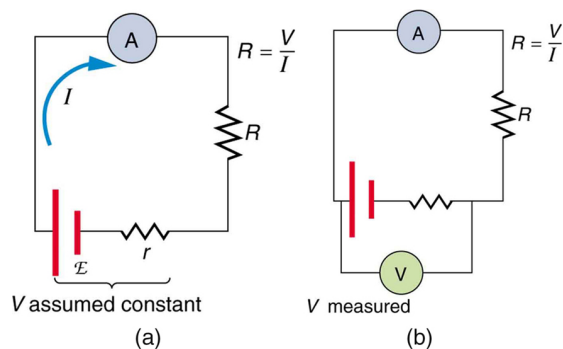
The potentiometer, a
null measurement

device. (a) A voltage source connected to a long wire resistor passes a constant current I through it. (b) An unknown emf (labeled script E_x in the figure) is connected as shown, and the point of contact along R is adjusted until the galvanometer reads zero. The segment of wire has a resistance R_x and script $E_x = IR_x$, where I is unaffected by the connection since no current flows through the galvanometer. The unknown emf is thus proportional to the resistance of the wire segment.

Because a long uniform wire is used for R , the ratio of resistances R_x/R_s is the same as the ratio of the lengths of wire that zero the galvanometer for each emf. The three quantities on the right-hand side of the equation are now known or measured, and emf_x can be calculated. The uncertainty in this calculation can be considerably smaller than when using a voltmeter directly, but it is not zero. There is always some uncertainty in the ratio of resistances R_x/R_s and in the standard emf_s . Furthermore, it is not possible to tell when the galvanometer reads exactly zero, which introduces error into both R_x and R_s , and may also affect the current I .

Resistance Measurements and the Wheatstone Bridge

There is a variety of so-called **ohmmeters** that purport to measure resistance. What the most common ohmmeters actually do is to apply a voltage to a resistance, measure the current, and calculate the resistance using Ohm's law. Their readout is this calculated resistance. Two configurations for ohmmeters using standard voltmeters and ammeters are shown in [\[link\]](#). Such configurations are limited in accuracy, because the meters alter both the voltage applied to the resistor and the current that flows through it.



Two methods for measuring resistance with standard meters. (a) Assuming a known voltage for the source, an ammeter measures current, and resistance is calculated as $R = \frac{V}{I}$. (b) Since the terminal voltage V varies with current, it is better to measure it. V is most accurately known when I is small, but I itself is most accurately known when it is large.

The **Wheatstone bridge** is a null measurement device for calculating resistance by balancing potential drops in a circuit. (See [\[link\]](#).) The device is called a bridge because the galvanometer forms a bridge between two branches. A variety of **bridge devices** are used to make null measurements in circuits.

Resistors R_1 and R_2 are precisely known, while the arrow through R_3 indicates that it is a variable resistance. The value of R_3 can be precisely read. With the unknown resistance R_x in the circuit, R_3 is adjusted until the galvanometer reads zero. The potential difference between points b and d is then zero, meaning that b and d are at the same potential. With no current running through the galvanometer, it has no effect on the rest of the circuit. So the branches abc and adc are in parallel, and each branch has the full voltage of the source. That is, the IR drops along abc and adc are the same. Since b and d are at the same potential, the IR drop along ad must equal the IR drop along ab. Thus,

Equation:

$$I_1 R_1 = I_2 R_3.$$

Again, since b and d are at the same potential, the IR drop along dc must equal the IR drop along bc. Thus,

Equation:

$$I_1 R_2 = I_2 R_x.$$

Taking the ratio of these last two expressions gives

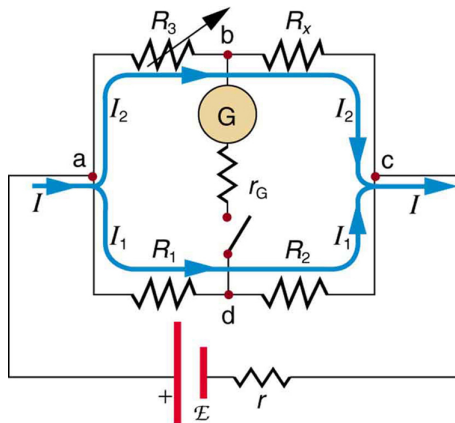
Equation:

$$\frac{I_1 R_1}{I_1 R_2} = \frac{I_2 R_3}{I_2 R_x}.$$

Canceling the currents and solving for R_x yields

Equation:

$$R_x = R_3 \frac{R_2}{R_1}.$$



The Wheatstone bridge is used to calculate unknown resistances. The variable resistance R_3 is adjusted until the galvanometer reads zero with the switch closed. This simplifies the circuit, allowing R_x to be calculated based on the IR drops as discussed in the text.

This equation is used to calculate the unknown resistance when current through the galvanometer is zero. This method can be very accurate (often to four significant digits), but it is limited by two factors. First, it is not possible to get the current through the galvanometer to be exactly zero.

Second, there are always uncertainties in R_1 , R_2 , and R_3 , which contribute to the uncertainty in R_x .

Exercise:

Check Your Understanding

Problem:

Identify other factors that might limit the accuracy of null measurements. Would the use of a digital device that is more sensitive than a galvanometer improve the accuracy of null measurements?

Solution:

One factor would be resistance in the wires and connections in a null measurement. These are impossible to make zero, and they can change over time. Another factor would be temperature variations in resistance, which can be reduced but not completely eliminated by choice of material. Digital devices sensitive to smaller currents than analog devices do improve the accuracy of null measurements because they allow you to get the current closer to zero.

Section Summary

- Null measurement techniques achieve greater accuracy by balancing a circuit so that no current flows through the measuring device.
- One such device, for determining voltage, is a potentiometer.
- Another null measurement device, for determining resistance, is the Wheatstone bridge.
- Other physical quantities can also be measured with null measurement techniques.

Conceptual questions

Exercise:

Problem:

Why can a null measurement be more accurate than one using standard voltmeters and ammeters? What factors limit the accuracy of null measurements?

Exercise:**Problem:**

If a potentiometer is used to measure cell emfs on the order of a few volts, why is it most accurate for the standard emf_s to be the same order of magnitude and the resistances to be in the range of a few ohms?

Problem Exercises**Exercise:****Problem:**

What is the emf_x of a cell being measured in a potentiometer, if the standard cell's emf is 12.0 V and the potentiometer balances for $R_x = 5.000\ \Omega$ and $R_s = 2.500\ \Omega$?

Solution:

24.0 V

Exercise:**Problem:**

Calculate the emf_x of a dry cell for which a potentiometer is balanced when $R_x = 1.200\ \Omega$, while an alkaline standard cell with an emf of 1.600 V requires $R_s = 1.247\ \Omega$ to balance the potentiometer.

Exercise:

Problem:

When an unknown resistance R_x is placed in a Wheatstone bridge, it is possible to balance the bridge by adjusting R_3 to be $2500\ \Omega$. What is R_x if $\frac{R_2}{R_1} = 0.625$?

Solution:

$1.56\ \text{k}\Omega$

Exercise:**Problem:**

To what value must you adjust R_3 to balance a Wheatstone bridge, if the unknown resistance R_x is $100\ \Omega$, R_1 is $50.0\ \Omega$, and R_2 is $175\ \Omega$?

Exercise:**Problem:**

(a) What is the unknown emf_x in a potentiometer that balances when R_x is $10.0\ \Omega$, and balances when R_s is $15.0\ \Omega$ for a standard 3.000-V emf ? (b) The same emf_x is placed in the same potentiometer, which now balances when R_s is $15.0\ \Omega$ for a standard emf of $3.100\ \text{V}$. At what resistance R_x will the potentiometer balance?

Solution:

(a) $2.00\ \text{V}$

(b) $9.68\ \Omega$

Exercise:**Problem:**

Suppose you want to measure resistances in the range from $10.0\ \Omega$ to $10.0\ \text{k}\Omega$ using a Wheatstone bridge that has $\frac{R_2}{R_1} = 2.000$. Over what range should R_3 be adjustable?

Solution:
Equation:

$$\text{Range} = 5.00 \, \Omega \text{ to } 5.00 \, \text{k}\Omega$$

Glossary

null measurements

methods of measuring current and voltage more accurately by balancing the circuit so that no current flows through the measurement device

potentiometer

a null measurement device for measuring potentials (voltages)

ohmmeter

an instrument that applies a voltage to a resistance, measures the current, calculates the resistance using Ohm's law, and provides a readout of this calculated resistance

bridge device

a device that forms a bridge between two branches of a circuit; some bridge devices are used to make null measurements in circuits

Wheatstone bridge

a null measurement device for calculating resistance by balancing potential drops in a circuit

Introduction to Magnetism

class="introduction"

The
magnificent
spectacle
of the
Aurora
Borealis, or
northern
lights,
glows in
the
northern
sky above
Bear Lake
near
Eielson Air
Force Base,
Alaska.
Shaped by
the Earth's
magnetic
field, this
light is
produced
by
radiation
spewed
from solar
storms.
(credit:
Senior
Airman
Joshua
Strang, via
Flickr)



One evening, an Alaskan sticks a note to his refrigerator with a small magnet. Through the kitchen window, the Aurora Borealis glows in the night sky. This grand spectacle is shaped by the same force that holds the note to the refrigerator.

People have been aware of magnets and magnetism for thousands of years. The earliest records date to well before the time of Christ, particularly in a region of Asia Minor called Magnesia (the name of this region is the source of words like *magnetic*). Magnetic rocks found in Magnesia, which is now part of western Turkey, stimulated interest during ancient times. A practical application for magnets was found later, when they were employed as navigational compasses. The use of magnets in compasses resulted not only in improved long-distance sailing, but also in the names of “north” and “south” being given to the two types of magnetic poles.

Today magnetism plays many important roles in our lives. Physicists’ understanding of magnetism has enabled the development of technologies that affect our everyday lives. The iPod in your purse or backpack, for example, wouldn’t have been possible without the applications of magnetism and electricity on a small scale.

The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology. The 2007 Nobel Prize in Physics went to Albert Fert from France and Peter Grunberg from Germany for this discovery of *giant magnetoresistance* and its applications to computer memory.

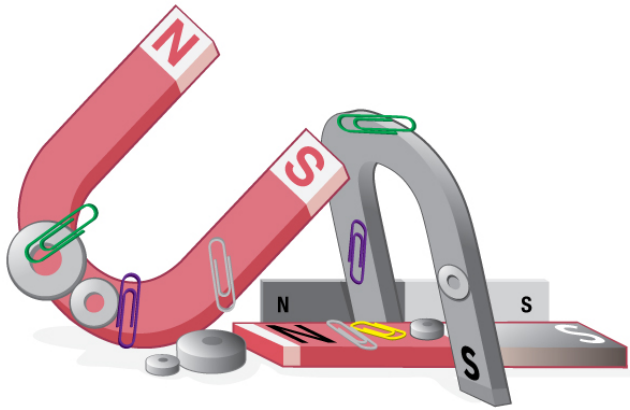
All electric motors, with uses as diverse as powering refrigerators, starting cars, and moving elevators, contain magnets. Generators, whether producing hydroelectric power or running bicycle lights, use magnetic fields. Recycling facilities employ magnets to separate iron from other refuse. Hundreds of millions of dollars are spent annually on magnetic containment of fusion as a future energy source. Magnetic resonance imaging (MRI) has become an important diagnostic tool in the field of medicine, and the use of magnetism to explore brain activity is a subject of contemporary research and development. The list of applications also includes computer hard drives, tape recording, detection of inhaled asbestos, and levitation of high-speed trains. Magnetism is used to explain atomic energy levels, cosmic rays, and charged particles trapped in the Van Allen belts. Once again, we will find all these disparate phenomena are linked by a small number of underlying physical principles.



Engineering of
technology like iPods
would not be possible
without a deep
understanding
magnetism. (credit: Jesse!
S?, Flickr)

Magnets

- Describe the difference between the north and south poles of a magnet.
- Describe how magnetic poles interact with each other.



Magnets come in various shapes, sizes, and strengths. All have both a north pole and a south pole. There is never an isolated pole (a monopole).

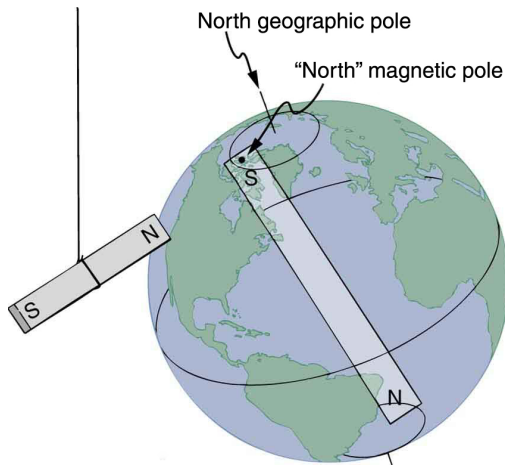
All magnets attract iron, such as that in a refrigerator door. However, magnets may attract or repel other magnets. Experimentation shows that all magnets have two poles. If freely suspended, one pole will point toward the north. The two poles are thus named the **north magnetic pole** and the **south magnetic pole** (or more properly, north-seeking and south-seeking poles, for the attractions in those directions).

Note:

Universal Characteristics of Magnets and Magnetic Poles

It is a universal characteristic of all magnets that *like poles repel and unlike poles attract*. (Note the similarity with electrostatics: unlike charges attract and like charges repel.)

Further experimentation shows that it is *impossible to separate north and south poles* in the manner that + and – charges can be separated.



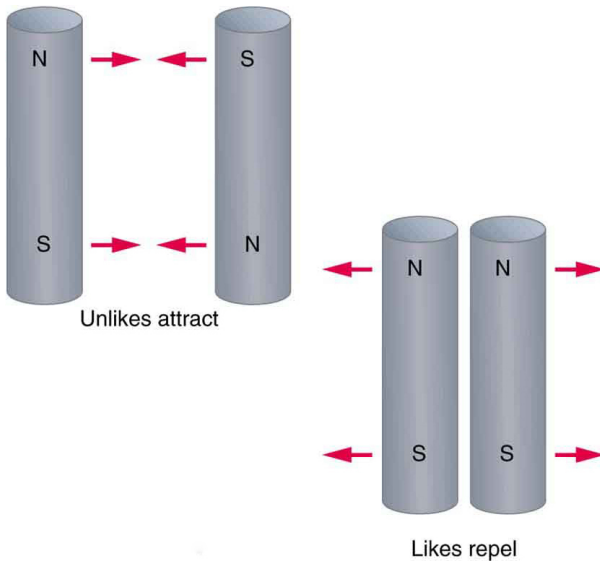
One end of a bar magnet is suspended from a thread that points toward north. The magnet's two poles are labeled N and S for north-seeking and south-seeking poles, respectively.

Note:

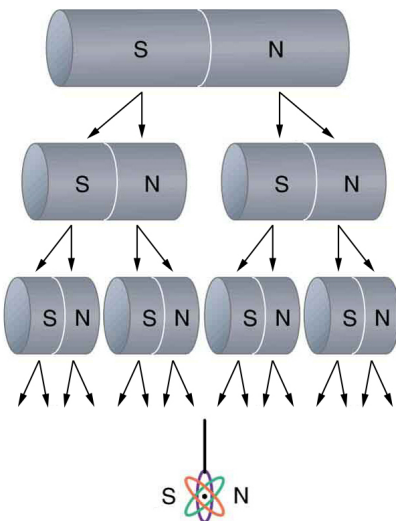
Misconception Alert: Earth's Geographic North Pole Hides an S

The Earth acts like a very large bar magnet with its south-seeking pole near the geographic North Pole. That is why the north pole of your compass is attracted toward the geographic north pole of the Earth—because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole! Confusion arises because the geographic term “North Pole” has come to be used (incorrectly) for the magnetic pole that is near the

North Pole. Thus, “North magnetic pole” is actually a misnomer—it should be called the South magnetic pole.



Unlike poles attract, whereas
like poles repel.



North and south
poles always occur
in pairs. Attempts

to separate them
result in more pairs
of poles. If we
continue to split the
magnet, we will
eventually get
down to an iron
atom with a north
pole and a south
pole—these, too,
cannot be
separated.

The fact that magnetic poles always occur in pairs of north and south is true from the very large scale—for example, sunspots always occur in pairs that are north and south magnetic poles—all the way down to the very small scale. Magnetic atoms have both a north pole and a south pole, as do many types of subatomic particles, such as electrons, protons, and neutrons.

Note:

Making Connections: Take-Home Experiment—Refrigerator Magnets

We know that like magnetic poles repel and unlike poles attract. See if you can show this for two refrigerator magnets. Will the magnets stick if you turn them over? Why do they stick to the door anyway? What can you say about the magnetic properties of the door next to the magnet? Do refrigerator magnets stick to metal or plastic spoons? Do they stick to all types of metal?

Section Summary

- Magnetism is a subject that includes the properties of magnets, the effect of the magnetic force on moving charges and currents, and the

- creation of magnetic fields by currents.
- There are two types of magnetic poles, called the north magnetic pole and south magnetic pole.
 - North magnetic poles are those that are attracted toward the Earth's geographic north pole.
 - Like poles repel and unlike poles attract.
 - Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.

Conceptual Questions

Exercise:

Problem:

Volcanic and other such activity at the mid-Atlantic ridge extrudes material to fill the gap between separating tectonic plates associated with continental drift. The magnetization of rocks is found to reverse in a coordinated manner with distance from the ridge. What does this imply about the Earth's magnetic field and how could the knowledge of the spreading rate be used to give its historical record?

Glossary

north magnetic pole

the end or the side of a magnet that is attracted toward Earth's geographic north pole

south magnetic pole

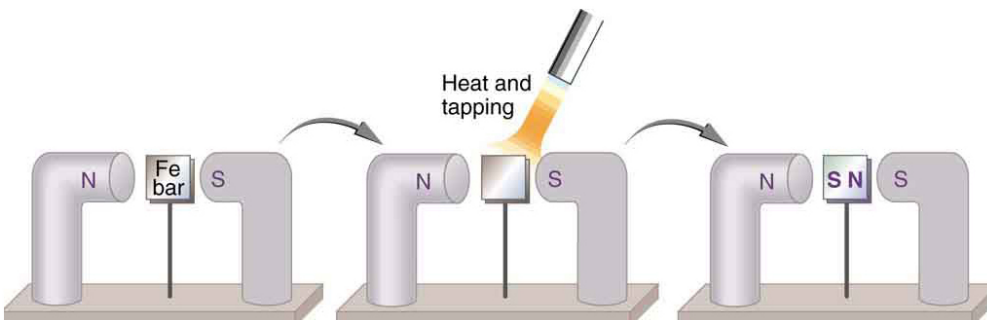
the end or the side of a magnet that is attracted toward Earth's geographic south pole

Ferromagnets and Electromagnets

- Define ferromagnet.
- Describe the role of magnetic domains in magnetization.
- Explain the significance of the Curie temperature.
- Describe the relationship between electricity and magnetism.

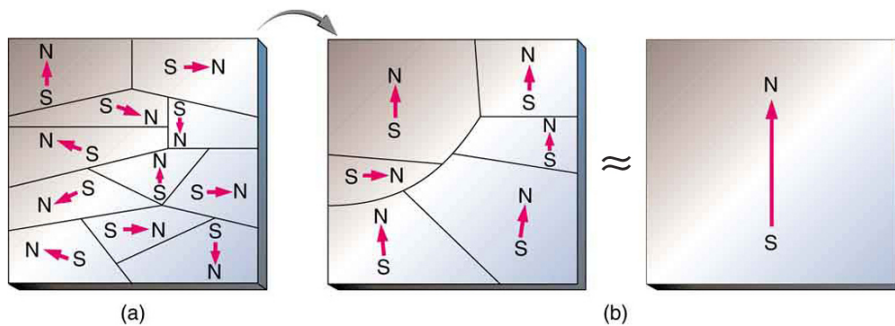
Ferromagnets

Only certain materials, such as iron, cobalt, nickel, and gadolinium, exhibit strong magnetic effects. Such materials are called **ferromagnetic**, after the Latin word for iron, *ferrum*. A group of materials made from the alloys of the rare earth elements are also used as strong and permanent magnets; a popular one is neodymium. Other materials exhibit weak magnetic effects, which are detectable only with sensitive instruments. Not only do ferromagnetic materials respond strongly to magnets (the way iron is attracted to magnets), they can also be **magnetized** themselves—that is, they can be induced to be magnetic or made into permanent magnets.



An unmagnetized piece of iron is placed between two magnets, heated, and then cooled, or simply tapped when cold. The iron becomes a permanent magnet with the poles aligned as shown: its south pole is adjacent to the north pole of the original magnet, and its north pole is adjacent to the south pole of the original magnet. Note that there are attractive forces between the magnets.

When a magnet is brought near a previously unmagnetized ferromagnetic material, it causes local magnetization of the material with unlike poles closest, as in [\[link\]](#). (This results in the attraction of the previously unmagnetized material to the magnet.) What happens on a microscopic scale is illustrated in [\[link\]](#). The regions within the material called **domains** act like small bar magnets. Within domains, the poles of individual atoms are aligned. Each atom acts like a tiny bar magnet. Domains are small and randomly oriented in an unmagnetized ferromagnetic object. In response to an external magnetic field, the domains may grow to millimeter size, aligning themselves as shown in [\[link\]](#)(b). This induced magnetization can be made permanent if the material is heated and then cooled, or simply tapped in the presence of other magnets.



(a) An unmagnetized piece of iron (or other ferromagnetic material) has randomly oriented domains. (b) When magnetized by an external field, the domains show greater alignment, and some grow at the expense of others. Individual atoms are aligned within domains; each atom acts like a tiny bar magnet.

Conversely, a permanent magnet can be demagnetized by hard blows or by heating it in the absence of another magnet. Increased thermal motion at higher temperature can disrupt and randomize the orientation and the size of

the domains. There is a well-defined temperature for ferromagnetic materials, which is called the **Curie temperature**, above which they cannot be magnetized. The Curie temperature for iron is 1043 K (770°C), which is well above room temperature. There are several elements and alloys that have Curie temperatures much lower than room temperature and are ferromagnetic only below those temperatures.

Electromagnets

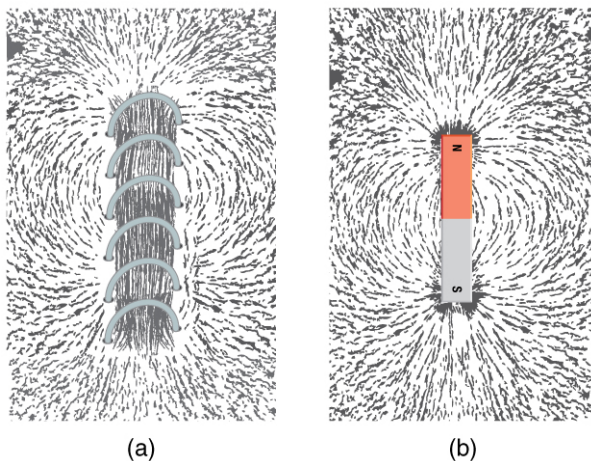
Early in the 19th century, it was discovered that electrical currents cause magnetic effects. The first significant observation was by the Danish scientist Hans Christian Oersted (1777–1851), who found that a compass needle was deflected by a current-carrying wire. This was the first significant evidence that the movement of charges had any connection with magnets. **Electromagnetism** is the use of electric current to make magnets. These temporarily induced magnets are called **electromagnets**. Electromagnets are employed for everything from a wrecking yard crane that lifts scrapped cars to controlling the beam of a 90-km-circumference particle accelerator to the magnets in medical imaging machines (See [\[link\]](#)).



Instrument for magnetic resonance imaging (MRI). The device uses a superconducting

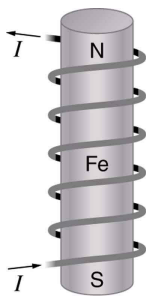
cylindrical coil for the main magnetic field. The patient goes into this “tunnel” on the gurney.
(credit: Bill McChesney, Flickr)

[\[link\]](#) shows that the response of iron filings to a current-carrying coil and to a permanent bar magnet. The patterns are similar. In fact, electromagnets and ferromagnets have the same basic characteristics—for example, they have north and south poles that cannot be separated and for which like poles repel and unlike poles attract.



Iron filings near (a) a current-carrying coil and (b) a magnet act like tiny compass needles, showing the shape of their fields. Their response to a current-carrying coil and a permanent magnet is seen to be very similar, especially near the ends of the coil and the magnet.

Combining a ferromagnet with an electromagnet can produce particularly strong magnetic effects. (See [\[link\]](#).) Whenever strong magnetic effects are needed, such as lifting scrap metal, or in particle accelerators, electromagnets are enhanced by ferromagnetic materials. Limits to how strong the magnets can be made are imposed by coil resistance (it will overheat and melt at sufficiently high current), and so superconducting magnets may be employed. These are still limited, because superconducting properties are destroyed by too great a magnetic field.

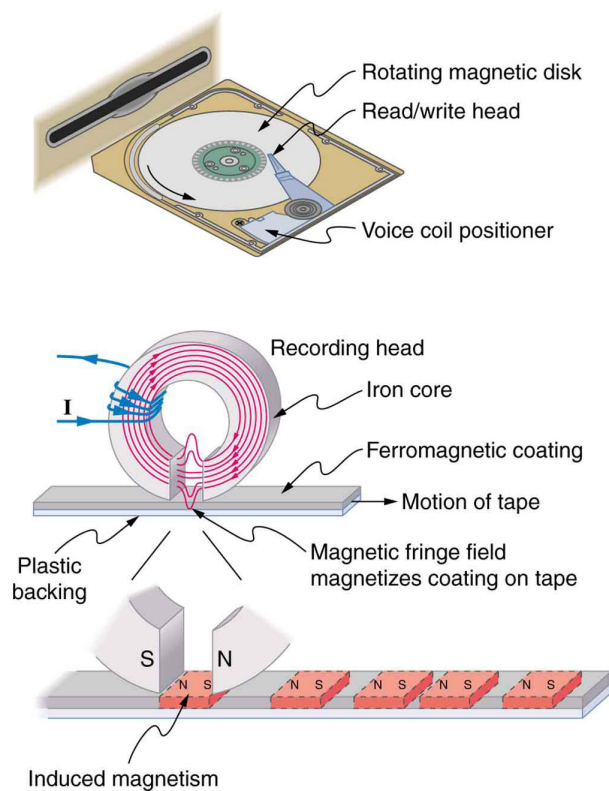


An
electromagnet
with a
ferromagnetic
core can
produce very
strong
magnetic
effects.

Alignment of
domains in the
core produces
a magnet, the
poles of which
are aligned
with the
electromagnet

.

[\[link\]](#) shows a few uses of combinations of electromagnets and ferromagnets. Ferromagnetic materials can act as memory devices, because the orientation of the magnetic fields of small domains can be reversed or erased. Magnetic information storage on videotapes and computer hard drives are among the most common applications. This property is vital in our digital world.



An electromagnet induces regions of permanent magnetism on a floppy disk coated with a ferromagnetic material. The information stored here is digital (a region is either magnetic or not); in other applications, it can be analog

(with a varying strength), such
as on audiotapes.

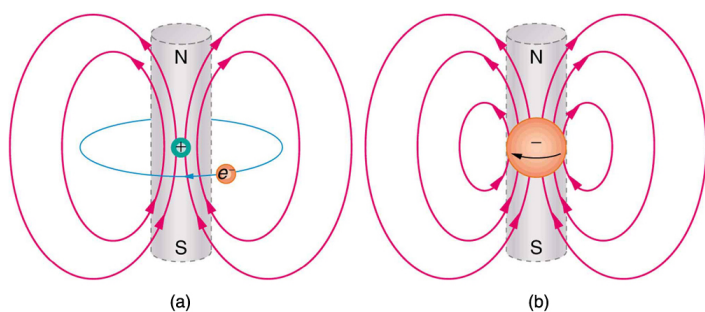
Current: The Source of All Magnetism

An electromagnet creates magnetism with an electric current. In later sections we explore this more quantitatively, finding the strength and direction of magnetic fields created by various currents. But what about ferromagnets? [\[link\]](#) shows models of how electric currents create magnetism at the submicroscopic level. (Note that we cannot directly observe the paths of individual electrons about atoms, and so a model or visual image, consistent with all direct observations, is made. We can directly observe the electron's orbital angular momentum, its spin momentum, and subsequent magnetic moments, all of which are explained with electric-current-creating subatomic magnetism.) Currents, including those associated with other submicroscopic particles like protons, allow us to explain ferromagnetism and all other magnetic effects. Ferromagnetism, for example, results from an internal cooperative alignment of electron spins, possible in some materials but not in others.

Crucial to the statement that electric current is the source of all magnetism is the fact that it is impossible to separate north and south magnetic poles. (This is far different from the case of positive and negative charges, which are easily separated.) A current loop always produces a magnetic dipole—that is, a magnetic field that acts like a north pole and south pole pair. Since isolated north and south magnetic poles, called **magnetic monopoles**, are not observed, currents are used to explain all magnetic effects. If magnetic monopoles did exist, then we would have to modify this underlying connection that all magnetism is due to electrical current. There is no known reason that magnetic monopoles should not exist—they are simply never observed—and so searches at the subnuclear level continue. If they *do not* exist, we would like to find out why not. If they *do* exist, we would like to see evidence of them.

Note:**Electric Currents and Magnetism**

Electric current is the source of all magnetism.



(a) In the planetary model of the atom, an electron orbits a nucleus, forming a closed-current loop and producing a magnetic field with a north pole and a south pole. (b) Electrons have spin and can be crudely pictured as rotating charge, forming a current that produces a magnetic field with a north pole and a south pole. Neither the planetary model nor the image of a spinning electron is completely consistent with modern physics. However, they do provide a useful way of understanding phenomena.

Note:**PhET Explorations: Magnets and Electromagnets**

Explore the interactions between a compass and bar magnet. Discover how you can use a battery and wire to make a magnet! Can you make it a stronger magnet? Can you make the magnetic field reverse?

Section Summary

- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.
- All magnetism is created by electric current.
- Ferromagnetic materials, such as iron, are those that exhibit strong magnetic effects.
- The atoms in ferromagnetic materials act like small magnets (due to currents within the atoms) and can be aligned, usually in millimeter-sized regions called domains.
- Domains can grow and align on a larger scale, producing permanent magnets. Such a material is magnetized, or induced to be magnetic.
- Above a material's Curie temperature, thermal agitation destroys the alignment of atoms, and ferromagnetism disappears.
- Electromagnets employ electric currents to make magnetic fields, often aided by induced fields in ferromagnetic materials.

Glossary

ferromagnetic

materials, such as iron, cobalt, nickel, and gadolinium, that exhibit strong magnetic effects

magnetized

to be turned into a magnet; to be induced to be magnetic

domains

regions within a material that behave like small bar magnets

Curie temperature

the temperature above which a ferromagnetic material cannot be magnetized

electromagnetism

the use of electrical currents to induce magnetism

electromagnet

an object that is temporarily magnetic when an electrical current is passed through it

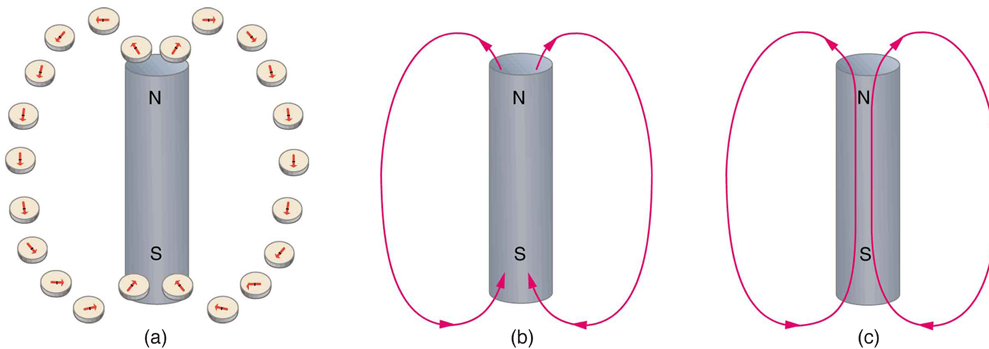
magnetic monopoles

an isolated magnetic pole; a south pole without a north pole, or vice versa (no magnetic monopole has ever been observed)

Magnetic Fields and Magnetic Field Lines

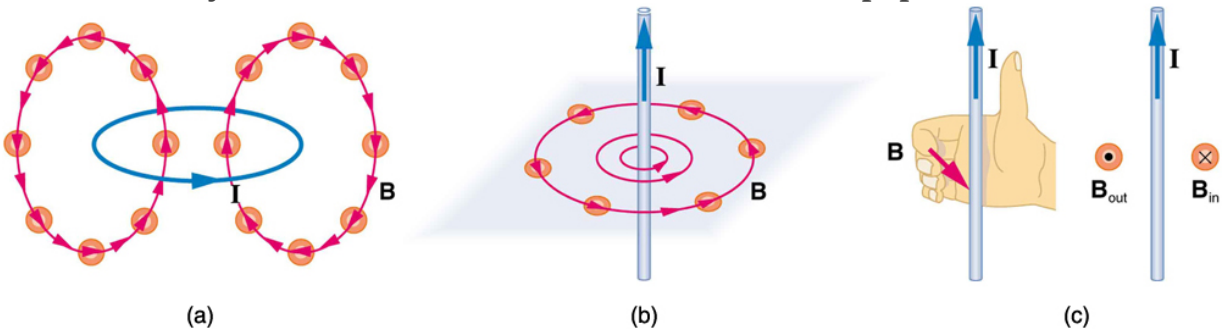
- Define magnetic field and describe the magnetic field lines of various magnetic fields.

Einstein is said to have been fascinated by a compass as a child, perhaps musing on how the needle felt a force without direct physical contact. His ability to think deeply and clearly about action at a distance, particularly for gravitational, electric, and magnetic forces, later enabled him to create his revolutionary theory of relativity. Since magnetic forces act at a distance, we define a **magnetic field** to represent magnetic forces. The pictorial representation of **magnetic field lines** is very useful in visualizing the strength and direction of the magnetic field. As shown in [\[link\]](#), the **direction of magnetic field lines** is defined to be the direction in which the north end of a compass needle points. The magnetic field is traditionally called the ***B*-field**.



Magnetic field lines are defined to have the direction that a small compass points when placed at a location. (a) If small compasses are used to map the magnetic field around a bar magnet, they will point in the directions shown: away from the north pole of the magnet, toward the south pole of the magnet. (Recall that the Earth's north magnetic pole is really a south pole in terms of definitions of poles on a bar magnet.) (b) Connecting the arrows gives continuous magnetic field lines. The strength of the field is proportional to the closeness (or density) of the lines. (c) If the interior of the magnet could be probed, the field lines would be found to form continuous closed loops.

Small compasses used to test a magnetic field will not disturb it. (This is analogous to the way we tested electric fields with a small test charge. In both cases, the fields represent only the object creating them and not the probe testing them.) [\[link\]](#) shows how the magnetic field appears for a current loop and a long straight wire, as could be explored with small compasses. A small compass placed in these fields will align itself parallel to the field line at its location, with its north pole pointing in the direction of B . Note the symbols used for field into and out of the paper.



Small compasses could be used to map the fields shown here. (a) The magnetic field of a circular current loop is similar to that of a bar magnet. (b) A long and straight wire creates a field with magnetic field lines forming circular loops. (c) When the wire is in the plane of the paper, the field is perpendicular to the paper. Note that the symbols used for the field pointing inward (like the tail of an arrow) and the field pointing outward (like the tip of an arrow).

Note:

Making Connections: Concept of a Field

A field is a way of mapping forces surrounding any object that can act on another object at a distance without apparent physical connection. The field represents the object generating it. Gravitational fields map

gravitational forces, electric fields map electrical forces, and magnetic fields map magnetic forces.

Extensive exploration of magnetic fields has revealed a number of hard-and-fast rules. We use magnetic field lines to represent the field (the lines are a pictorial tool, not a physical entity in and of themselves). The properties of magnetic field lines can be summarized by these rules:

1. The direction of the magnetic field is tangent to the field line at any point in space. A small compass will point in the direction of the field line.
2. The strength of the field is proportional to the closeness of the lines. It is exactly proportional to the number of lines per unit area perpendicular to the lines (called the areal density).
3. Magnetic field lines can never cross, meaning that the field is unique at any point in space.
4. Magnetic field lines are continuous, forming closed loops without beginning or end. They go from the north pole to the south pole.

The last property is related to the fact that the north and south poles cannot be separated. It is a distinct difference from electric field lines, which begin and end on the positive and negative charges. If magnetic monopoles existed, then magnetic field lines would begin and end on them.

Section Summary

- Magnetic fields can be pictorially represented by magnetic field lines, the properties of which are as follows:
 1. The field is tangent to the magnetic field line.
 2. Field strength is proportional to the line density.
 3. Field lines cannot cross.
 4. Field lines are continuous loops.

Conceptual Questions

Exercise:**Problem:**

Explain why the magnetic field would not be unique (that is, not have a single value) at a point in space where magnetic field lines might cross. (Consider the direction of the field at such a point.)

Exercise:**Problem:**

List the ways in which magnetic field lines and electric field lines are similar. For example, the field direction is tangent to the line at any point in space. Also list the ways in which they differ. For example, electric force is parallel to electric field lines, whereas magnetic force on moving charges is perpendicular to magnetic field lines.

Exercise:**Problem:**

Noting that the magnetic field lines of a bar magnet resemble the electric field lines of a pair of equal and opposite charges, do you expect the magnetic field to rapidly decrease in strength with distance from the magnet? Is this consistent with your experience with magnets?

Exercise:**Problem:**

Is the Earth's magnetic field parallel to the ground at all locations? If not, where is it parallel to the surface? Is its strength the same at all locations? If not, where is it greatest?

Glossary

magnetic field

the representation of magnetic forces

B-field

another term for magnetic field

magnetic field lines

the pictorial representation of the strength and the direction of a magnetic field

direction of magnetic field lines

the direction that the north end of a compass needle points

Magnetic Field Strength: Force on a Moving Charge in a Magnetic Field

- Describe the effects of magnetic fields on moving charges.
- Use the right hand rule 1 to determine the velocity of a charge, the direction of the magnetic field, and the direction of the magnetic force on a moving charge.
- Calculate the magnetic force on a moving charge.

What is the mechanism by which one magnet exerts a force on another?

The answer is related to the fact that all magnetism is caused by current, the flow of charge. *Magnetic fields exert forces on moving charges*, and so they exert forces on other magnets, all of which have moving charges.

Right Hand Rule 1

The magnetic force on a moving charge is one of the most fundamental known. Magnetic force is as important as the electrostatic or Coulomb force. Yet the magnetic force is more complex, in both the number of factors that affects it and in its direction, than the relatively simple Coulomb force. The magnitude of the **magnetic force** F on a charge q moving at a speed v in a magnetic field of strength B is given by

Equation:

$$F = qvB \sin \theta,$$

where θ is the angle between the directions of \mathbf{v} and \mathbf{B} . This force is often called the **Lorentz force**. In fact, this is how we define the magnetic field strength B —in terms of the force on a charged particle moving in a magnetic field. The SI unit for magnetic field strength B is called the **tesla** (T) after the eccentric but brilliant inventor Nikola Tesla (1856–1943). To determine how the tesla relates to other SI units, we solve $F = qvB \sin \theta$ for B .

Equation:

$$B = \frac{F}{qv \sin \theta}$$

Because $\sin \theta$ is unitless, the tesla is

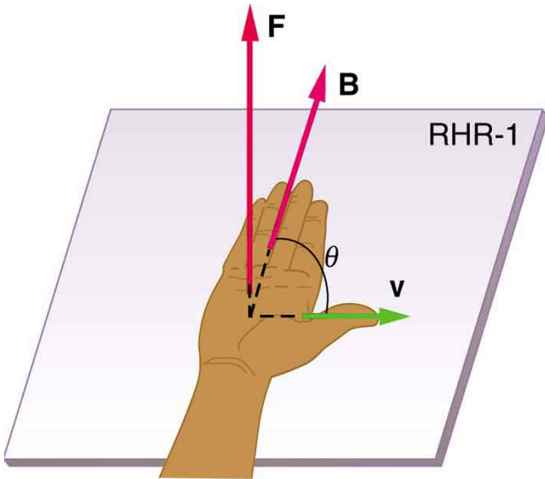
Equation:

$$1 \text{ T} = \frac{1 \text{ N}}{\text{C} \cdot \text{m/s}} = \frac{1 \text{ N}}{\text{A} \cdot \text{m}}$$

(note that $\text{C/s} = \text{A}$).

Another smaller unit, called the **gauss** (G), where $1 \text{ G} = 10^{-4} \text{ T}$, is sometimes used. The strongest permanent magnets have fields near 2 T; superconducting electromagnets may attain 10 T or more. The Earth's magnetic field on its surface is only about $5 \times 10^{-5} \text{ T}$, or 0.5 G.

The *direction* of the magnetic force **F** is perpendicular to the plane formed by **v** and **B**, as determined by the **right hand rule 1** (or RHR-1), which is illustrated in [\[link\]](#). RHR-1 states that, to determine the direction of the magnetic force on a positive moving charge, you point the thumb of the right hand in the direction of **v**, the fingers in the direction of **B**, and a perpendicular to the palm points in the direction of **F**. One way to remember this is that there is one velocity, and so the thumb represents it. There are many field lines, and so the fingers represent them. The force is in the direction you would push with your palm. The force on a negative charge is in exactly the opposite direction to that on a positive charge.



$$F = qvB \sin \theta$$

$\mathbf{F} \perp \text{plane of } \mathbf{v} \text{ and } \mathbf{B}$

Magnetic fields exert forces on moving charges. This force is one of the most basic known. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by \mathbf{v} and \mathbf{B} and follows right hand rule—1 (RHR-1) as shown. The magnitude of the force is proportional to q , v , B , and the sine of the angle between \mathbf{v} and \mathbf{B} .

Note:

Making Connections: Charges and Magnets

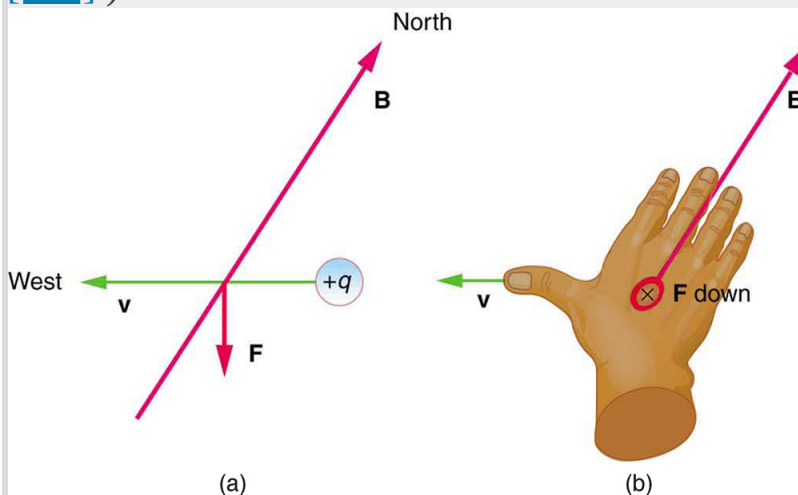
There is no magnetic force on static charges. However, there is a magnetic force on moving charges. When charges are stationary, their electric fields do not affect magnets. But, when charges move, they produce magnetic

fields that exert forces on other magnets. When there is relative motion, a connection between electric and magnetic fields emerges—each affects the other.

Example:

Calculating Magnetic Force: Earth's Magnetic Field on a Charged Glass Rod

With the exception of compasses, you seldom see or personally experience forces due to the Earth's small magnetic field. To illustrate this, suppose that in a physics lab you rub a glass rod with silk, placing a 20-nC positive charge on it. Calculate the force on the rod due to the Earth's magnetic field, if you throw it with a horizontal velocity of 10 m/s due west in a place where the Earth's field is due north parallel to the ground. (The direction of the force is determined with right hand rule 1 as shown in [\[link\]](#).)



A positively charged object moving due west in a region where the Earth's magnetic field is due north experiences a force that is straight down as shown. A negative charge moving in the same direction would feel a force straight up.

Strategy

We are given the charge, its velocity, and the magnetic field strength and direction. We can thus use the equation $F = qvB \sin \theta$ to find the force.

Solution

The magnetic force is

Equation:

$$F = qvb \sin \theta.$$

We see that $\sin \theta = 1$, since the angle between the velocity and the direction of the field is 90° . Entering the other given quantities yields

Equation:

$$\begin{aligned} F &= (20 \times 10^{-9} \text{ C})(10 \text{ m/s})(5 \times 10^{-5} \text{ T}) \\ &= 1 \times 10^{-11} (\text{C} \cdot \text{m/s}) \left(\frac{\text{N}}{\text{C} \cdot \text{m/s}} \right) = 1 \times 10^{-11} \text{ N}. \end{aligned}$$

Discussion

This force is completely negligible on any macroscopic object, consistent with experience. (It is calculated to only one digit, since the Earth's field varies with location and is given to only one digit.) The Earth's magnetic field, however, does produce very important effects, particularly on submicroscopic particles. Some of these are explored in [Force on a Moving Charge in a Magnetic Field: Examples and Applications](#).

Section Summary

- Magnetic fields exert a force on a moving charge q , the magnitude of which is

Equation:

$$F = qvB \sin \theta,$$

where θ is the angle between the directions of v and B .

- The SI unit for magnetic field strength B is the tesla (T), which is related to other units by

Equation:

$$1 \text{ T} = \frac{1 \text{ N}}{\text{C} \cdot \text{m/s}} = \frac{1 \text{ N}}{\text{A} \cdot \text{m}}.$$

- The *direction* of the force on a moving charge is given by right hand rule 1 (RHR-1): Point the thumb of the right hand in the direction of v , the fingers in the direction of B , and a perpendicular to the palm points in the direction of F .
- The force is perpendicular to the plane formed by \mathbf{v} and \mathbf{B} . Since the force is zero if \mathbf{v} is parallel to \mathbf{B} , charged particles often follow magnetic field lines rather than cross them.

Conceptual Questions

Exercise:

Problem:

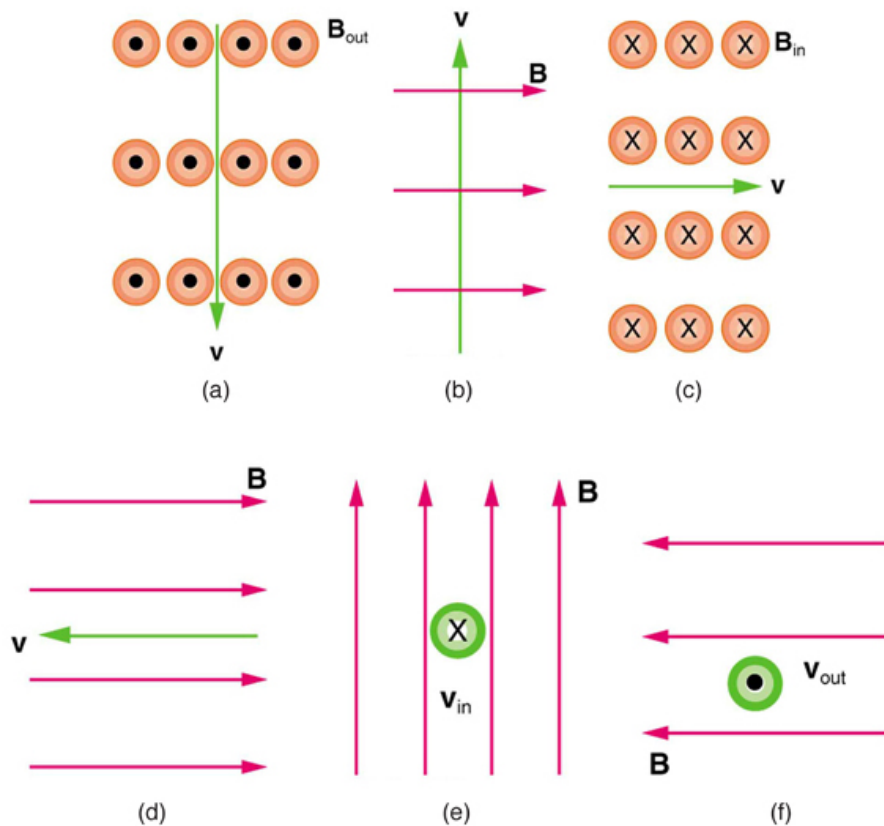
If a charged particle moves in a straight line through some region of space, can you say that the magnetic field in that region is necessarily zero?

Problems & Exercises

Exercise:

Problem:

What is the direction of the magnetic force on a positive charge that moves as shown in each of the six cases shown in [\[link\]](#)?



Solution:

- (a) Left (West)
- (b) Into the page
- (c) Up (North)
- (d) No force
- (e) Right (East)
- (f) Down (South)

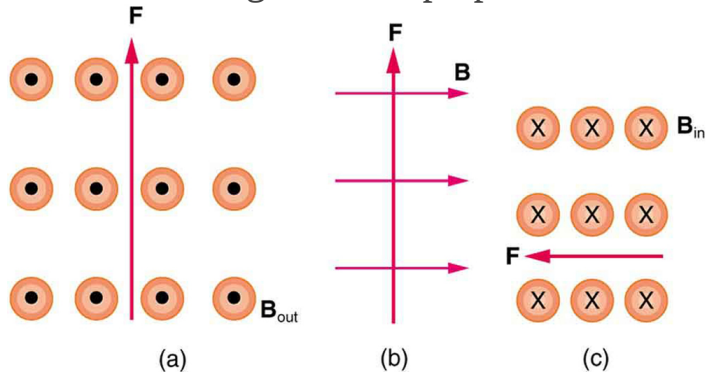
Exercise:

Problem: Repeat [\[link\]](#) for a negative charge.

Exercise:

Problem:

What is the direction of the velocity of a negative charge that experiences the magnetic force shown in each of the three cases in [\[link\]](#), assuming it moves perpendicular to \mathbf{B} ?

**Solution:**

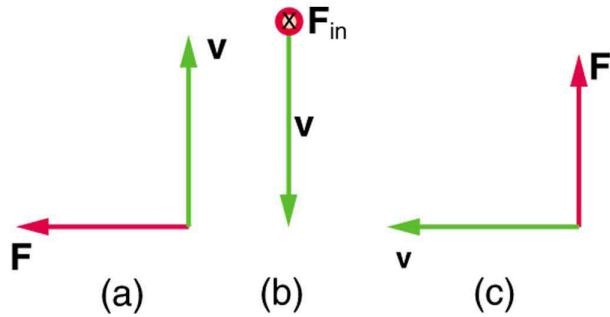
- (a) East (right)
- (b) Into page
- (c) South (down)

Exercise:

Problem: Repeat [\[link\]](#) for a positive charge.

Exercise:**Problem:**

What is the direction of the magnetic field that produces the magnetic force on a positive charge as shown in each of the three cases in the figure below, assuming \mathbf{B} is perpendicular to \mathbf{v} ?



Solution:

(a) Into page

(b) West (left)

(c) Out of page

Exercise:

Problem: Repeat [\[link\]](#) for a negative charge.

Exercise:

Problem:

What is the maximum magnitude of the force on an aluminum rod with a $0.100\text{-}\mu\text{C}$ charge that you pass between the poles of a 1.50-T permanent magnet at a speed of 5.00 m/s ? In what direction is the force?

Solution:

$7.50 \times 10^{-7}\text{ N}$ perpendicular to both the magnetic field lines and the velocity

Exercise:

Problem:

(a) Aircraft sometimes acquire small static charges. Suppose a supersonic jet has a $0.500\text{-}\mu\text{C}$ charge and flies due west at a speed of 660 m/s over the Earth's magnetic south pole (near Earth's geographic north pole), where the $8.00 \times 10^{-5}\text{-T}$ magnetic field points straight down. What are the direction and the magnitude of the magnetic force on the plane? (b) Discuss whether the value obtained in part (a) implies this is a significant or negligible effect.

Exercise:**Problem:**

(a) A cosmic ray proton moving toward the Earth at $5.00 \times 10^7\text{ m/s}$ experiences a magnetic force of $1.70 \times 10^{-16}\text{ N}$. What is the strength of the magnetic field if there is a 45° angle between it and the proton's velocity? (b) Is the value obtained in part (a) consistent with the known strength of the Earth's magnetic field on its surface? Discuss.

Solution:

(a) $3.01 \times 10^{-5}\text{ T}$

(b) This is slightly less than the magnetic field strength of $5 \times 10^{-5}\text{ T}$ at the surface of the Earth, so it is consistent.

Exercise:**Problem:**

An electron moving at $4.00 \times 10^3\text{ m/s}$ in a 1.25-T magnetic field experiences a magnetic force of $1.40 \times 10^{-16}\text{ N}$. What angle does the velocity of the electron make with the magnetic field? There are two answers.

Exercise:

Problem:

(a) A physicist performing a sensitive measurement wants to limit the magnetic force on a moving charge in her equipment to less than 1.00×10^{-12} N. What is the greatest the charge can be if it moves at a maximum speed of 30.0 m/s in the Earth's field? (b) Discuss whether it would be difficult to limit the charge to less than the value found in (a) by comparing it with typical static electricity and noting that static is often absent.

Solution:

(a) 6.67×10^{-10} C (taking the Earth's field to be 5.00×10^{-5} T)

(b) Less than typical static, therefore difficult

Glossary

right hand rule 1 (RHR-1)

the rule to determine the direction of the magnetic force on a positive moving charge: when the thumb of the right hand points in the direction of the charge's velocity \mathbf{v} and the fingers point in the direction of the magnetic field \mathbf{B} , then the force on the charge is perpendicular and away from the palm; the force on a negative charge is perpendicular and into the palm

Lorentz force

the force on a charge moving in a magnetic field

tesla

T, the SI unit of the magnetic field strength; $1 \text{ T} = \frac{1 \text{ N}}{\text{A} \cdot \text{m}}$

magnetic force

the force on a charge produced by its motion through a magnetic field;
the Lorentz force

gauss

G, the unit of the magnetic field strength; $1 \text{ G} = 10^{-4} \text{ T}$

Force on a Moving Charge in a Magnetic Field: Examples and Applications

- Describe the effects of a magnetic field on a moving charge.
- Calculate the radius of curvature of the path of a charge that is moving in a magnetic field.

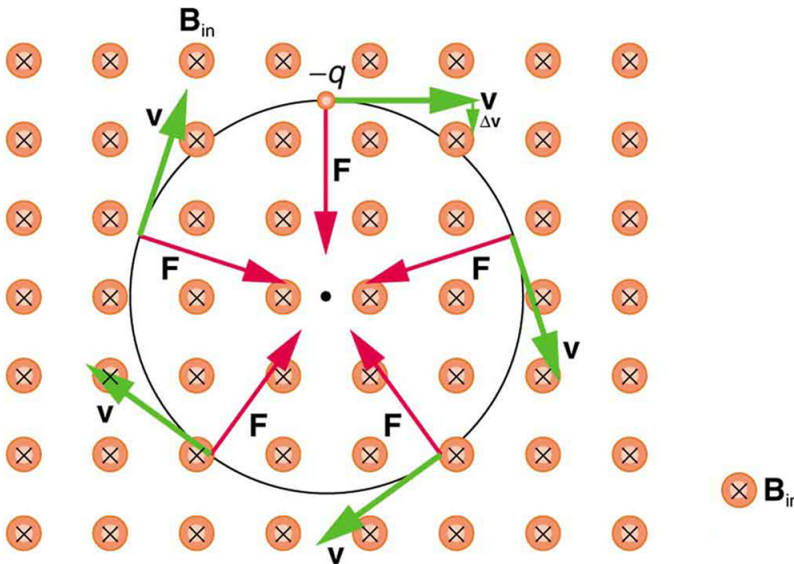
Magnetic force can cause a charged particle to move in a circular or spiral path. Cosmic rays are energetic charged particles in outer space, some of which approach the Earth. They can be forced into spiral paths by the Earth's magnetic field. Protons in giant accelerators are kept in a circular path by magnetic force. The bubble chamber photograph in [\[link\]](#) shows charged particles moving in such curved paths. The curved paths of charged particles in magnetic fields are the basis of a number of phenomena and can even be used analytically, such as in a mass spectrometer.



Trails of bubbles are produced by high-energy charged particles moving through the superheated liquid hydrogen in this artist's rendition of a bubble chamber. There is a strong magnetic field perpendicular to the page that causes the curved paths of the particles. The radius of the path can be

used to find the mass,
charge, and energy of the
particle.

So does the magnetic force cause circular motion? Magnetic force is always perpendicular to velocity, so that it does no work on the charged particle. The particle's kinetic energy and speed thus remain constant. The direction of motion is affected, but not the speed. This is typical of uniform circular motion. The simplest case occurs when a charged particle moves perpendicular to a uniform B -field, such as shown in [\[link\]](#). (If this takes place in a vacuum, the magnetic field is the dominant factor determining the motion.) Here, the magnetic force supplies the centripetal force $F_c = mv^2/r$. Noting that $\sin \theta = 1$, we see that $F = qvB$.



A negatively charged particle moves in the plane of the page in a region where the magnetic field is perpendicular into the page (represented by the small circles with x's—like the tails of arrows). The magnetic force is perpendicular to the velocity, and so velocity changes in direction but not magnitude. Uniform circular motion results.

Because the magnetic force F supplies the centripetal force F_c , we have
Equation:

$$qvB = \frac{mv^2}{r}.$$

Solving for r yields
Equation:

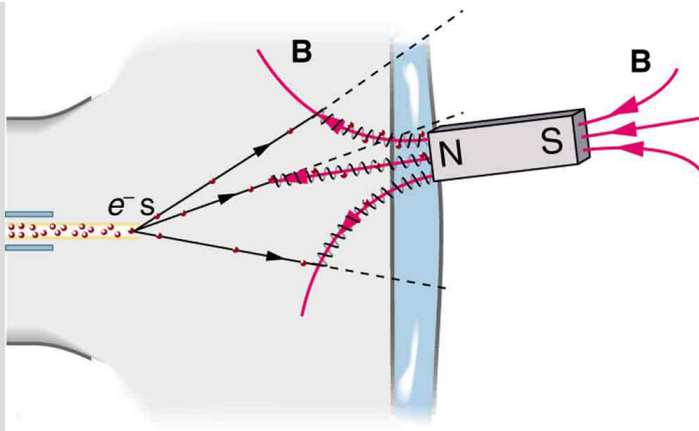
$$r = \frac{mv}{qB}.$$

Here, r is the radius of curvature of the path of a charged particle with mass m and charge q , moving at a speed v perpendicular to a magnetic field of strength B . If the velocity is not perpendicular to the magnetic field, then v is the component of the velocity perpendicular to the field. The component of the velocity parallel to the field is unaffected, since the magnetic force is zero for motion parallel to the field. This produces a spiral motion rather than a circular one.

Example:

Calculating the Curvature of the Path of an Electron Moving in a Magnetic Field: A Magnet on a TV Screen

A magnet brought near an old-fashioned TV screen such as in [\[link\]](#) (TV sets with cathode ray tubes instead of LCD screens) severely distorts its picture by altering the path of the electrons that make its phosphors glow. ***(Don't try this at home, as it will permanently magnetize and ruin the TV.)*** To illustrate this, calculate the radius of curvature of the path of an electron having a velocity of 6.00×10^7 m/s (corresponding to the accelerating voltage of about 10.0 kV used in some TVs) perpendicular to a magnetic field of strength $B = 0.500$ T (obtainable with permanent magnets).



Side view showing what happens when a magnet comes in contact with a computer monitor or TV screen. Electrons moving toward the screen spiral about magnetic field lines, maintaining the component of their velocity parallel to the field lines. This distorts the image on the screen.

Strategy

We can find the radius of curvature r directly from the equation $r = \frac{mv}{qB}$, since all other quantities in it are given or known.

Solution

Using known values for the mass and charge of an electron, along with the given values of v and B gives us

Equation:

$$\begin{aligned} r = \frac{mv}{qB} &= \frac{(9.11 \times 10^{-31} \text{ kg})(6.00 \times 10^7 \text{ m/s})}{(1.60 \times 10^{-19} \text{ C})(0.500 \text{ T})} \\ &= 6.83 \times 10^{-4} \text{ m} \end{aligned}$$

or

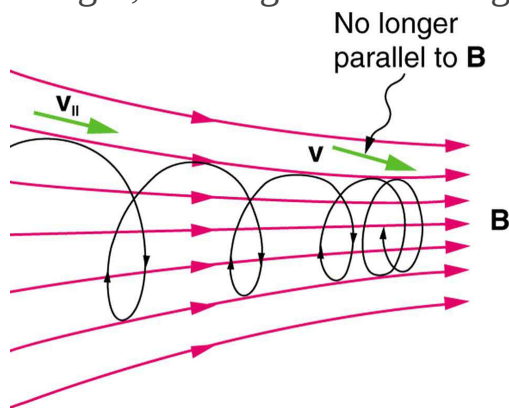
Equation:

$$r = 0.683 \text{ mm.}$$

Discussion

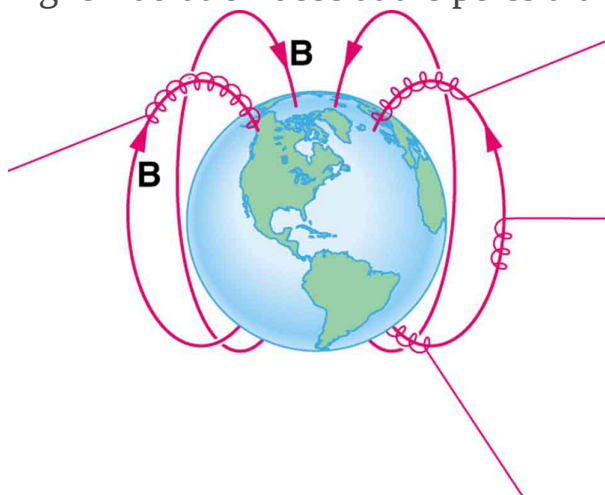
The small radius indicates a large effect. The electrons in the TV picture tube are made to move in very tight circles, greatly altering their paths and distorting the image.

[\[link\]](#) shows how electrons not moving perpendicular to magnetic field lines follow the field lines. The component of velocity parallel to the lines is unaffected, and so the charges spiral along the field lines. If field strength increases in the direction of motion, the field will exert a force to slow the charges, forming a kind of magnetic mirror, as shown below.



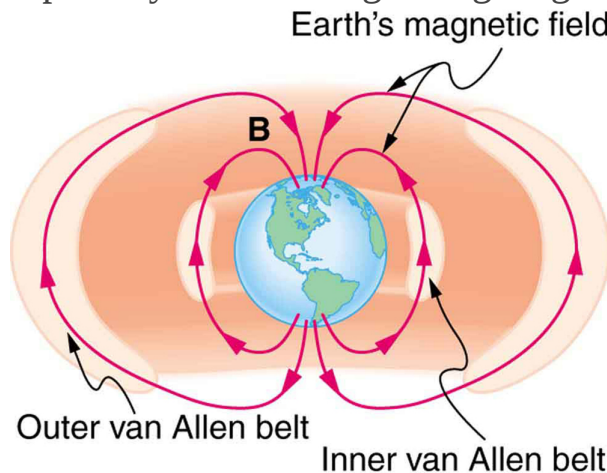
When a charged particle moves along a magnetic field line into a region where the field becomes stronger, the particle experiences a force that reduces the component of velocity parallel to the field. This force slows the motion along the field line and here reverses it, forming a “magnetic mirror.”

The properties of charged particles in magnetic fields are related to such different things as the Aurora Australis or Aurora Borealis and particle accelerators. *Charged particles approaching magnetic field lines may get trapped in spiral orbits about the lines rather than crossing them*, as seen above. Some cosmic rays, for example, follow the Earth's magnetic field lines, entering the atmosphere near the magnetic poles and causing the southern or northern lights through their ionization of molecules in the atmosphere. This glow of energized atoms and molecules is seen in [\[link\]](#). Those particles that approach middle latitudes must cross magnetic field lines, and many are prevented from penetrating the atmosphere. Cosmic rays are a component of background radiation; consequently, they give a higher radiation dose at the poles than at the equator.



Energetic electrons and protons, components of cosmic rays, from the Sun and deep outer space often follow the Earth's magnetic field lines rather than cross them. (Recall that the Earth's north magnetic pole is really a south pole in terms of a bar magnet.)

Some incoming charged particles become trapped in the Earth's magnetic field, forming two belts above the atmosphere known as the Van Allen radiation belts after the discoverer James A. Van Allen, an American astrophysicist. (See [\[link\]](#).) Particles trapped in these belts form radiation fields (similar to nuclear radiation) so intense that manned space flights avoid them and satellites with sensitive electronics are kept out of them. In the few minutes it took lunar missions to cross the Van Allen radiation belts, astronauts received radiation doses more than twice the allowed annual exposure for radiation workers. Other planets have similar belts, especially those having strong magnetic fields like Jupiter.



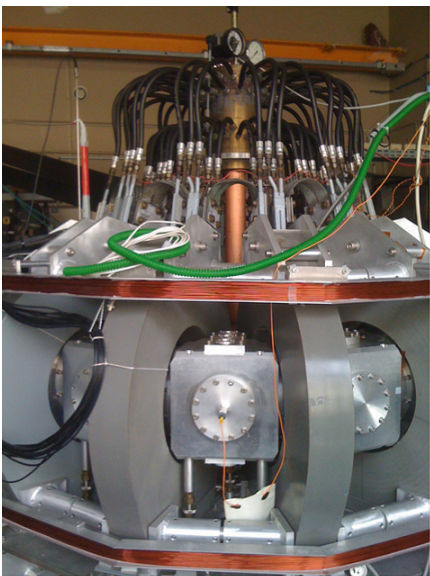
The Van Allen radiation belts are two regions in which energetic charged particles are trapped in the Earth's magnetic field. One belt lies about 300 km above the Earth's surface, the other about 16,000 km. Charged particles in these belts migrate along magnetic field lines and are partially reflected away from the poles by the stronger fields there. The charged particles that enter the atmosphere are replenished by the Sun and sources in deep outer space.

Back on Earth, we have devices that employ magnetic fields to contain charged particles. Among them are the giant particle accelerators that have been used to explore the substructure of matter. (See [\[link\]](#).) Magnetic fields not only control the direction of the charged particles, they also are used to focus particles into beams and overcome the repulsion of like charges in these beams.

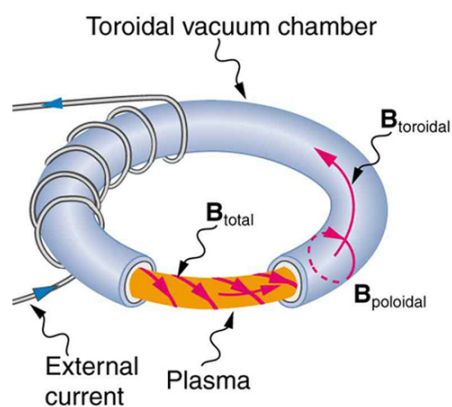


The Fermilab facility in Illinois has a large particle accelerator (the most powerful in the world until 2008) that employs magnetic fields (magnets seen here in orange) to contain and direct its beam. This and other accelerators have been in use for several decades and have allowed us to discover some of the laws underlying all matter. (credit: ammcgrim, Flickr)

Thermonuclear fusion (like that occurring in the Sun) is a hope for a future clean energy source. One of the most promising devices is the *tokamak*, which uses magnetic fields to contain (or trap) and direct the reactive charged particles. (See [\[link\]](#).) Less exotic, but more immediately practical, amplifiers in microwave ovens use a magnetic field to contain oscillating electrons. These oscillating electrons generate the microwaves sent into the oven.



(a)



(b)

Tokamaks such as the one shown in the figure are being studied with the goal of economical production of energy by nuclear fusion. Magnetic fields in the doughnut-shaped device contain and direct the reactive charged particles. (credit: David Mellis, Flickr)

Mass spectrometers have a variety of designs, and many use magnetic fields to measure mass. The curvature of a charged particle's path in the field is related to its mass and is measured to obtain mass information. (See [More Applications of Magnetism](#).) Historically, such techniques were employed in the first direct observations of electron charge and mass. Today, mass

spectrometers (sometimes coupled with gas chromatographs) are used to determine the make-up and sequencing of large biological molecules.

Section Summary

- Magnetic force can supply centripetal force and cause a charged particle to move in a circular path of radius

Equation:

$$r = \frac{mv}{qB},$$

where v is the component of the velocity perpendicular to B for a charged particle with mass m and charge q .

Conceptual Questions

Exercise:

Problem:

How can the motion of a charged particle be used to distinguish between a magnetic and an electric field?

Exercise:

Problem:

High-velocity charged particles can damage biological cells and are a component of radiation exposure in a variety of locations ranging from research facilities to natural background. Describe how you could use a magnetic field to shield yourself.

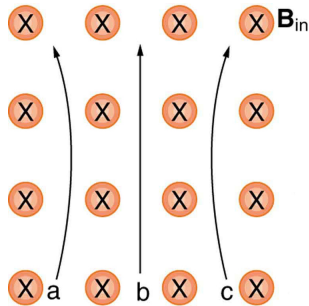
Exercise:

Problem:

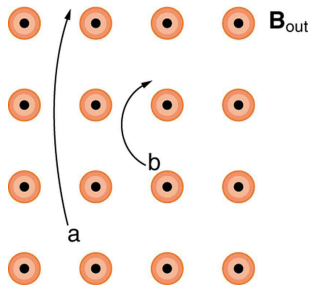
If a cosmic ray proton approaches the Earth from outer space along a line toward the center of the Earth that lies in the plane of the equator, in what direction will it be deflected by the Earth's magnetic field? What about an electron? A neutron?

Exercise:

Problem: What are the signs of the charges on the particles in [\[link\]](#)?

**Exercise:****Problem:**

Which of the particles in [\[link\]](#) has the greatest velocity, assuming they have identical charges and masses?

**Exercise:****Problem:**

Which of the particles in [\[link\]](#) has the greatest mass, assuming all have identical charges and velocities?

Exercise:

Problem:

While operating, a high-precision TV monitor is placed on its side during maintenance. The image on the monitor changes color and blurs slightly. Discuss the possible relation of these effects to the Earth's magnetic field.

Problems & Exercises

If you need additional support for these problems, see [More Applications of Magnetism](#).

Exercise:**Problem:**

A cosmic ray electron moves at 7.50×10^6 m/s perpendicular to the Earth's magnetic field at an altitude where field strength is 1.00×10^{-5} T. What is the radius of the circular path the electron follows?

Solution:

4.27 m

Exercise:**Problem:**

A proton moves at 7.50×10^7 m/s perpendicular to a magnetic field. The field causes the proton to travel in a circular path of radius 0.800 m. What is the field strength?

Exercise:

Problem:

(a) Viewers of *Star Trek* hear of an antimatter drive on the Starship *Enterprise*. One possibility for such a futuristic energy source is to store antimatter charged particles in a vacuum chamber, circulating in a magnetic field, and then extract them as needed. Antimatter annihilates with normal matter, producing pure energy. What strength magnetic field is needed to hold antiprotons, moving at 5.00×10^7 m/s in a circular path 2.00 m in radius? Antiprotons have the same mass as protons but the opposite (negative) charge. (b) Is this field strength obtainable with today's technology or is it a futuristic possibility?

Solution:

(a) 0.261 T

(b) This strength is definitely obtainable with today's technology. Magnetic field strengths of 0.500 T are obtainable with permanent magnets.

Exercise:**Problem:**

(a) An oxygen-16 ion with a mass of 2.66×10^{-26} kg travels at 5.00×10^6 m/s perpendicular to a 1.20-T magnetic field, which makes it move in a circular arc with a 0.231-m radius. What positive charge is on the ion? (b) What is the ratio of this charge to the charge of an electron? (c) Discuss why the ratio found in (b) should be an integer.

Exercise:**Problem:**

What radius circular path does an electron travel if it moves at the same speed and in the same magnetic field as the proton in [\[link\]](#)?

Solution:

$$4.36 \times 10^{-4} \text{ m}$$

Exercise:**Problem:**

A velocity selector in a mass spectrometer uses a 0.100-T magnetic field. (a) What electric field strength is needed to select a speed of $4.00 \times 10^6 \text{ m/s}$? (b) What is the voltage between the plates if they are separated by 1.00 cm?

Exercise:**Problem:**

An electron in a TV CRT moves with a speed of $6.00 \times 10^7 \text{ m/s}$, in a direction perpendicular to the Earth's field, which has a strength of $5.00 \times 10^{-5} \text{ T}$. (a) What strength electric field must be applied perpendicular to the Earth's field to make the electron moves in a straight line? (b) If this is done between plates separated by 1.00 cm, what is the voltage applied? (Note that TVs are usually surrounded by a ferromagnetic material to shield against external magnetic fields and avoid the need for such a correction.)

Solution:

(a) 3.00 kV/m

(b) 30.0 V

Exercise:

Problem:

(a) At what speed will a proton move in a circular path of the same radius as the electron in [\[link\]](#)? (b) What would the radius of the path be if the proton had the same speed as the electron? (c) What would the radius be if the proton had the same kinetic energy as the electron? (d) The same momentum?

Exercise:**Problem:**

A mass spectrometer is being used to separate common oxygen-16 from the much rarer oxygen-18, taken from a sample of old glacial ice. (The relative abundance of these oxygen isotopes is related to climatic temperature at the time the ice was deposited.) The ratio of the masses of these two ions is 16 to 18, the mass of oxygen-16 is 2.66×10^{-26} kg, and they are singly charged and travel at 5.00×10^6 m/s in a 1.20-T magnetic field. What is the separation between their paths when they hit a target after traversing a semicircle?

Solution:

0.173 m

Exercise:**Problem:**

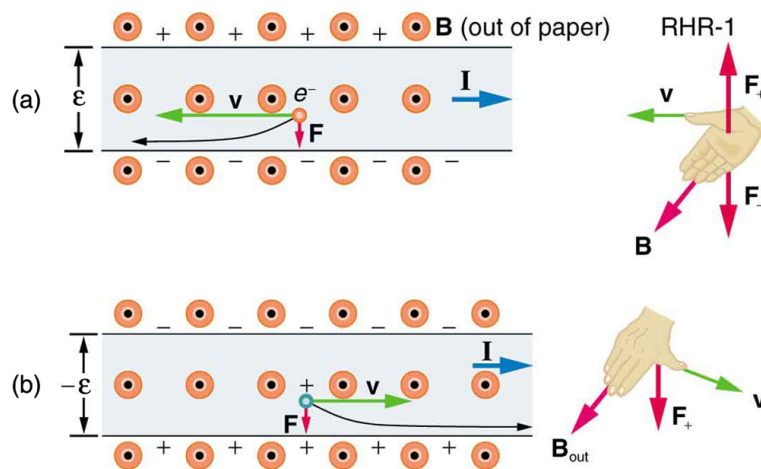
(a) Triply charged uranium-235 and uranium-238 ions are being separated in a mass spectrometer. (The much rarer uranium-235 is used as reactor fuel.) The masses of the ions are 3.90×10^{-25} kg and 3.95×10^{-25} kg, respectively, and they travel at 3.00×10^5 m/s in a 0.250-T field. What is the separation between their paths when they hit a target after traversing a semicircle? (b) Discuss whether this distance between their paths seems to be big enough to be practical in the separation of uranium-235 from uranium-238.

The Hall Effect

- Describe the Hall effect.
- Calculate the Hall emf across a current-carrying conductor.

We have seen effects of a magnetic field on free-moving charges. The magnetic field also affects charges moving in a conductor. One result is the Hall effect, which has important implications and applications.

[\[link\]](#) shows what happens to charges moving through a conductor in a magnetic field. The field is perpendicular to the electron drift velocity and to the width of the conductor. Note that conventional current is to the right in both parts of the figure. In part (a), electrons carry the current and move to the left. In part (b), positive charges carry the current and move to the right. Moving electrons feel a magnetic force toward one side of the conductor, leaving a net positive charge on the other side. This separation of charge *creates a voltage* ε , known as the **Hall emf**, *across* the conductor. The creation of a voltage *across* a current-carrying conductor by a magnetic field is known as the **Hall effect**, after Edwin Hall, the American physicist who discovered it in 1879.



The Hall effect. (a) Electrons move to the left in this flat conductor (conventional current to the right). The magnetic field is directly out of the page, represented by circled dots; it exerts a force on the moving charges, causing a voltage ε , the

Hall emf, across the conductor. (b)
Positive charges moving to the right
(conventional current also to the right) are
moved to the side, producing a Hall emf
of the opposite sign, $-\varepsilon$. Thus, if the
direction of the field and current are
known, the sign of the charge carriers can
be determined from the Hall effect.

One very important use of the Hall effect is to determine whether positive or negative charges carries the current. Note that in [\[link\]](#)(b), where positive charges carry the current, the Hall emf has the sign opposite to when negative charges carry the current. Historically, the Hall effect was used to show that electrons carry current in metals and it also shows that positive charges carry current in some semiconductors. The Hall effect is used today as a research tool to probe the movement of charges, their drift velocities and densities, and so on, in materials. In 1980, it was discovered that the Hall effect is quantized, an example of quantum behavior in a macroscopic object.

The Hall effect has other uses that range from the determination of blood flow rate to precision measurement of magnetic field strength. To examine these quantitatively, we need an expression for the Hall emf, ε , across a conductor. Consider the balance of forces on a moving charge in a situation where B , v , and l are mutually perpendicular, such as shown in [\[link\]](#). Although the magnetic force moves negative charges to one side, they cannot build up without limit. The electric field caused by their separation opposes the magnetic force, $F = qvB$, and the electric force, $F_e = qE$, eventually grows to equal it. That is,

Equation:

$$qE = qvB$$

or

Equation:

$$E = vB.$$

Note that the electric field E is uniform across the conductor because the magnetic field B is uniform, as is the conductor. For a uniform electric field, the relationship between electric field and voltage is $E = \varepsilon/l$, where l is the width of the conductor and ε is the Hall emf. Entering this into the last expression gives

Equation:

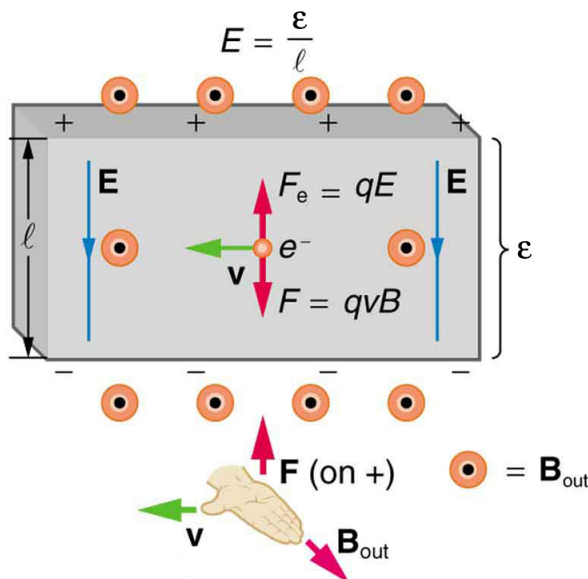
$$\frac{\varepsilon}{l} = vB.$$

Solving this for the Hall emf yields

Equation:

$$\varepsilon = Blv \text{ (} B, v, \text{ and } l, \text{ mutually perpendicular),}$$

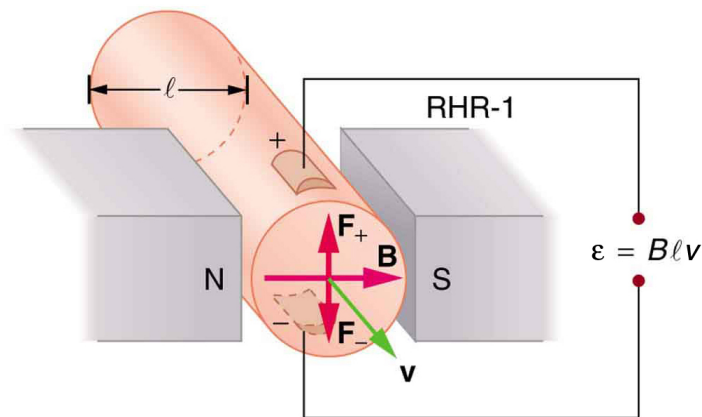
where ε is the Hall effect voltage across a conductor of width l through which charges move at a speed v .



The Hall emf ε produces an electric force that balances the magnetic force on the moving

charges. The magnetic force produces charge separation, which builds up until it is balanced by the electric force, an equilibrium that is quickly reached.

One of the most common uses of the Hall effect is in the measurement of magnetic field strength B . Such devices, called *Hall probes*, can be made very small, allowing fine position mapping. Hall probes can also be made very accurate, usually accomplished by careful calibration. Another application of the Hall effect is to measure fluid flow in any fluid that has free charges (most do). (See [\[link\]](#).) A magnetic field applied perpendicular to the flow direction produces a Hall emf ε as shown. Note that the sign of ε depends not on the sign of the charges, but only on the directions of B and v . The magnitude of the Hall emf is $\varepsilon = Blv$, where l is the pipe diameter, so that the average velocity v can be determined from ε providing the other factors are known.



The Hall effect can be used to measure fluid flow in any fluid having free charges, such as blood. The Hall emf ε is measured across the tube perpendicular to the applied magnetic field and is proportional to the average velocity v .

Example:**Calculating the Hall emf: Hall Effect for Blood Flow**

A Hall effect flow probe is placed on an artery, applying a 0.100-T magnetic field across it, in a setup similar to that in [\[link\]](#). What is the Hall emf, given the vessel's inside diameter is 4.00 mm and the average blood velocity is 20.0 cm/s?

Strategy

Because B , v , and l are mutually perpendicular, the equation $\varepsilon = Blv$ can be used to find ε .

Solution

Entering the given values for B , v , and l gives

Equation:

$$\begin{aligned}\varepsilon &= Blv = (0.100 \text{ T})(4.00 \times 10^{-3} \text{ m})(0.200 \text{ m/s}) \\ &= 80.0 \text{ } \mu\text{V}\end{aligned}$$

Discussion

This is the average voltage output. Instantaneous voltage varies with pulsating blood flow. The voltage is small in this type of measurement. ε is particularly difficult to measure, because there are voltages associated with heart action (ECG voltages) that are on the order of millivolts. In practice, this difficulty is overcome by applying an AC magnetic field, so that the Hall emf is AC with the same frequency. An amplifier can be very selective in picking out only the appropriate frequency, eliminating signals and noise at other frequencies.

Section Summary

- The Hall effect is the creation of voltage ε , known as the Hall emf, across a current-carrying conductor by a magnetic field.
- The Hall emf is given by

Equation:

$$\varepsilon = Blv \text{ (} B, v, \text{ and } l, \text{ mutually perpendicular)}$$

for a conductor of width l through which charges move at a speed v .

Conceptual Questions

Exercise:

Problem:

Discuss how the Hall effect could be used to obtain information on free charge density in a conductor. (Hint: Consider how drift velocity and current are related.)

Problems & Exercises

Exercise:

Problem:

A large water main is 2.50 m in diameter and the average water velocity is 6.00 m/s. Find the Hall voltage produced if the pipe runs perpendicular to the Earth's 5.00×10^{-5} -T field.

Solution:

$$7.50 \times 10^{-4} \text{ V}$$

Exercise:

Problem:

What Hall voltage is produced by a 0.200-T field applied across a 2.60-cm-diameter aorta when blood velocity is 60.0 cm/s?

Exercise:

Problem:

(a) What is the speed of a supersonic aircraft with a 17.0-m wingspan, if it experiences a 1.60-V Hall voltage between its wing tips when in level flight over the north magnetic pole, where the Earth's field strength is $8.00 \times 10^{-5} \text{ T}$? (b) Explain why very little current flows as a result of this Hall voltage.

Solution:

(a) $1.18 \times 10^3 \text{ m/s}$

(b) Once established, the Hall emf pushes charges one direction and the magnetic force acts in the opposite direction resulting in no net force on the charges. Therefore, no current flows in the direction of the Hall emf. This is the same as in a current-carrying conductor—current does not flow in the direction of the Hall emf.

Exercise:**Problem:**

A nonmechanical water meter could utilize the Hall effect by applying a magnetic field across a metal pipe and measuring the Hall voltage produced. What is the average fluid velocity in a 3.00-cm-diameter pipe, if a 0.500-T field across it creates a 60.0-mV Hall voltage?

Exercise:**Problem:**

Calculate the Hall voltage induced on a patient's heart while being scanned by an MRI unit. Approximate the conducting path on the heart wall by a wire 7.50 cm long that moves at 10.0 cm/s perpendicular to a 1.50-T magnetic field.

Solution:

11.3 mV

Exercise:**Problem:**

A Hall probe calibrated to read $1.00\ \mu\text{V}$ when placed in a 2.00-T field is placed in a 0.150-T field. What is its output voltage?

Exercise:**Problem:**

Using information in [\[link\]](#), what would the Hall voltage be if a 2.00-T field is applied across a 10-gauge copper wire (2.588 mm in diameter) carrying a 20.0-A current?

Solution:

$1.16\ \mu\text{V}$

Exercise:**Problem:**

Show that the Hall voltage across wires made of the same material, carrying identical currents, and subjected to the same magnetic field is inversely proportional to their diameters. (Hint: Consider how drift velocity depends on wire diameter.)

Exercise:**Problem:**

A patient with a pacemaker is mistakenly being scanned for an MRI image. A 10.0-cm -long section of pacemaker wire moves at a speed of 10.0 cm/s perpendicular to the MRI unit's magnetic field and a 20.0-mV Hall voltage is induced. What is the magnetic field strength?

Solution:

2.00 T

Glossary

Hall effect

the creation of voltage across a current-carrying conductor by a magnetic field

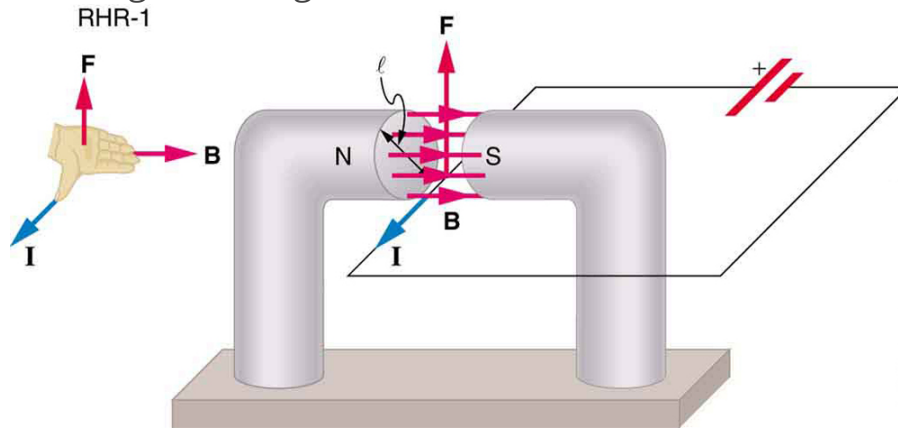
Hall emf

the electromotive force created by a current-carrying conductor by a magnetic field, $\varepsilon = Blv$

Magnetic Force on a Current-Carrying Conductor

- Describe the effects of a magnetic force on a current-carrying conductor.
- Calculate the magnetic force on a current-carrying conductor.

Because charges ordinarily cannot escape a conductor, the magnetic force on charges moving in a conductor is transmitted to the conductor itself.



The magnetic field exerts a force on a current-carrying wire in a direction given by the right hand rule 1 (the same direction as that on the individual moving charges). This force can easily be large enough to move the wire, since typical currents consist of very large numbers of moving charges.

We can derive an expression for the magnetic force on a current by taking a sum of the magnetic forces on individual charges. (The forces add because they are in the same direction.) The force on an individual charge moving at the drift velocity v_d is given by $F = qv_d B \sin \theta$. Taking B to be uniform over a length of wire l and zero elsewhere, the total magnetic force on the wire is then $F = (qv_d B \sin \theta)(N)$, where N is the number of charge carriers in the section of wire of length l . Now, $N = nV$, where n is the number of charge carriers per unit volume and V is the volume of wire in the field. Noting that $V = Al$, where A is the cross-sectional area of the

wire, then the force on the wire is $F = (qv_d B \sin \theta)(nAl)$. Gathering terms,

Equation:

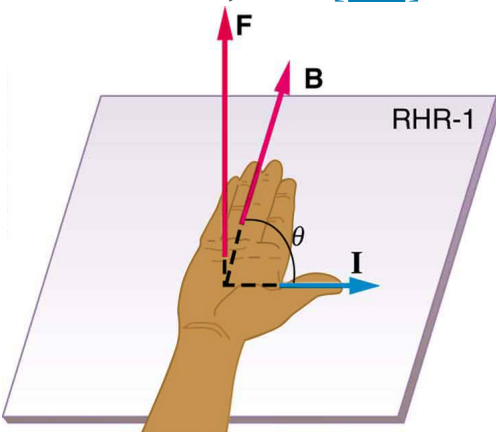
$$F = (nqAv_d)lB \sin \theta.$$

Because $nqAv_d = I$ (see [Current](#)),

Equation:

$$F = IlB \sin \theta$$

is the equation for *magnetic force on a length l of wire carrying a current I in a uniform magnetic field B* , as shown in [\[link\]](#). If we divide both sides of this expression by l , we find that the magnetic force per unit length of wire in a uniform field is $\frac{F}{l} = IB \sin \theta$. The direction of this force is given by RHR-1, with the thumb in the direction of the current I . Then, with the fingers in the direction of B , a perpendicular to the palm points in the direction of F , as in [\[link\]](#).



$$F = IlB \sin \theta$$

$\mathbf{F} \perp$ plane of \mathbf{I} and \mathbf{B}

The force on a current-carrying wire in a magnetic field is $F = IlB \sin \theta$. Its

direction is given by
RHR-1.

Example:

Calculating Magnetic Force on a Current-Carrying Wire: A Strong Magnetic Field

Calculate the force on the wire shown in [\[link\]](#), given $B = 1.50 \text{ T}$, $l = 5.00 \text{ cm}$, and $I = 20.0 \text{ A}$.

Strategy

The force can be found with the given information by using $F = IlB \sin \theta$ and noting that the angle θ between I and B is 90° , so that $\sin \theta = 1$.

Solution

Entering the given values into $F = IlB \sin \theta$ yields

Equation:

$$F = IlB \sin \theta = (20.0 \text{ A})(0.0500 \text{ m})(1.50 \text{ T})(1).$$

The units for tesla are $1 \text{ T} = \frac{\text{N}}{\text{A}\cdot\text{m}}$; thus,

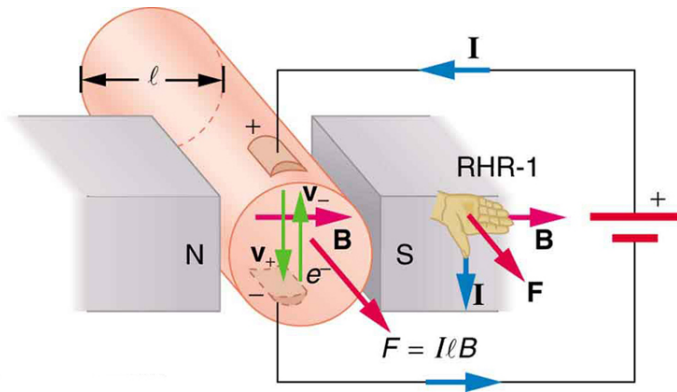
Equation:

$$F = 1.50 \text{ N}.$$

Discussion

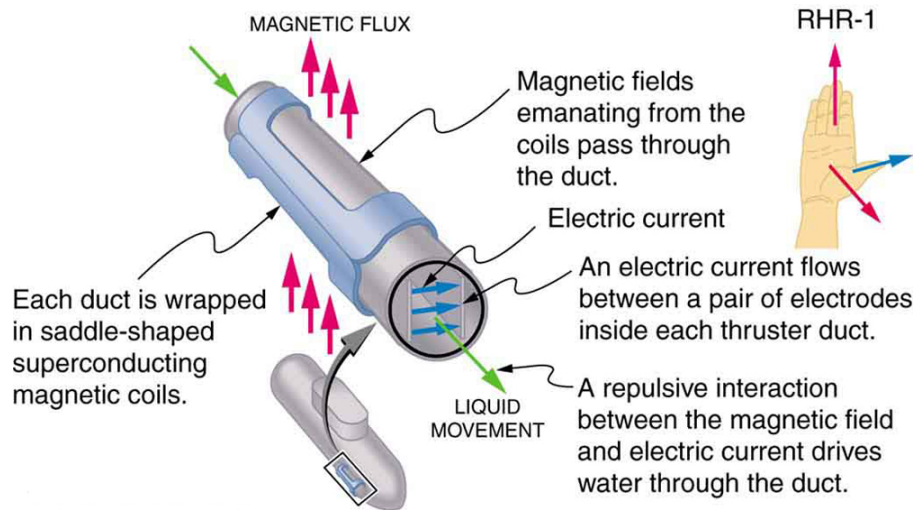
This large magnetic field creates a significant force on a small length of wire.

Magnetic force on current-carrying conductors is used to convert electric energy to work. (Motors are a prime example—they employ loops of wire and are considered in the next section.) Magnetohydrodynamics (MHD) is the technical name given to a clever application where magnetic force pumps fluids without moving mechanical parts. (See [\[link\]](#).)



Magnetohydrodynamics. The magnetic force on the current passed through this fluid can be used as a nonmechanical pump.

A strong magnetic field is applied across a tube and a current is passed through the fluid at right angles to the field, resulting in a force on the fluid parallel to the tube axis as shown. The absence of moving parts makes this attractive for moving a hot, chemically active substance, such as the liquid sodium employed in some nuclear reactors. Experimental artificial hearts are testing with this technique for pumping blood, perhaps circumventing the adverse effects of mechanical pumps. (Cell membranes, however, are affected by the large fields needed in MHD, delaying its practical application in humans.) MHD propulsion for nuclear submarines has been proposed, because it could be considerably quieter than conventional propeller drives. The deterrent value of nuclear submarines is based on their ability to hide and survive a first or second nuclear strike. As we slowly disassemble our nuclear weapons arsenals, the submarine branch will be the last to be decommissioned because of this ability (See [\[link\]](#).) Existing MHD drives are heavy and inefficient—much development work is needed.



An MHD propulsion system in a nuclear submarine could produce significantly less turbulence than propellers and allow it to run more silently. The development of a silent drive submarine was dramatized in the book and the film *The Hunt for Red October*.

Section Summary

- The magnetic force on current-carrying conductors is given by **Equation:**

$$F = IlB \sin \theta,$$

where I is the current, l is the length of a straight conductor in a uniform magnetic field B , and θ is the angle between I and B . The force follows RHR-1 with the thumb in the direction of I .

Conceptual Questions

Exercise:

Problem:

Draw a sketch of the situation in [\[link\]](#) showing the direction of electrons carrying the current, and use RHR-1 to verify the direction of the force on the wire.

Exercise:**Problem:**

Verify that the direction of the force in an MHD drive, such as that in [\[link\]](#), does not depend on the sign of the charges carrying the current across the fluid.

Exercise:**Problem:**

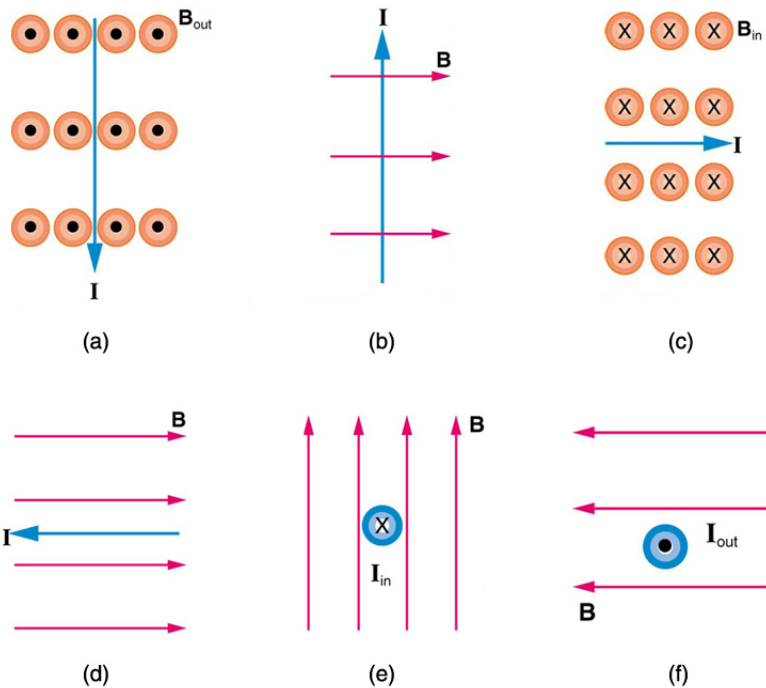
Why would a magnetohydrodynamic drive work better in ocean water than in fresh water? Also, why would superconducting magnets be desirable?

Exercise:**Problem:**

Which is more likely to interfere with compass readings, AC current in your refrigerator or DC current when you start your car? Explain.

Problems & Exercises**Exercise:****Problem:**

What is the direction of the magnetic force on the current in each of the six cases in [\[link\]](#)?



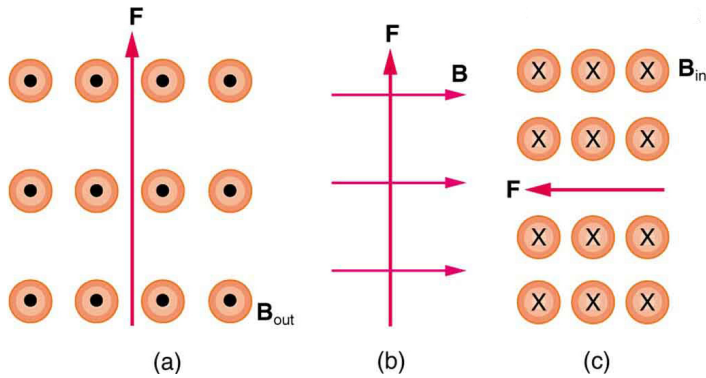
Solution:

- (a) west (left)
- (b) into page
- (c) north (up)
- (d) no force
- (e) east (right)
- (f) south (down)

Exercise:

Problem:

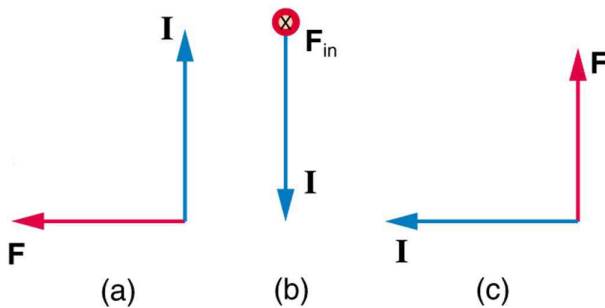
What is the direction of a current that experiences the magnetic force shown in each of the three cases in [\[link\]](#), assuming the current runs perpendicular to B ?



Exercise:

Problem:

What is the direction of the magnetic field that produces the magnetic force shown on the currents in each of the three cases in [\[link\]](#), assuming \mathbf{I} is perpendicular to \mathbf{B} ?



Solution:

(a) into page

(b) west (left)

(c) out of page

Exercise:

Problem:

(a) What is the force per meter on a lightning bolt at the equator that carries 20,000 A perpendicular to the Earth's 3.00×10^{-5} -T field? (b) What is the direction of the force if the current is straight up and the Earth's field direction is due north, parallel to the ground?

Exercise:**Problem:**

(a) A DC power line for a light-rail system carries 1000 A at an angle of 30.0° to the Earth's 5.00×10^{-5} -T field. What is the force on a 100-m section of this line? (b) Discuss practical concerns this presents, if any.

Solution:

(a) 2.50 N

(b) This is about half a pound of force per 100 m of wire, which is much less than the weight of the wire itself. Therefore, it does not cause any special concerns.

Exercise:**Problem:**

What force is exerted on the water in an MHD drive utilizing a 25.0-cm-diameter tube, if 100-A current is passed across the tube that is perpendicular to a 2.00-T magnetic field? (The relatively small size of this force indicates the need for very large currents and magnetic fields to make practical MHD drives.)

Exercise:**Problem:**

A wire carrying a 30.0-A current passes between the poles of a strong magnet that is perpendicular to its field and experiences a 2.16-N force on the 4.00 cm of wire in the field. What is the average field strength?

Solution:

1.80 T

Exercise:

Problem:

(a) A 0.750-m-long section of cable carrying current to a car starter motor makes an angle of 60° with the Earth's 5.50×10^{-5} T field. What is the current when the wire experiences a force of 7.00×10^{-3} N? (b) If you run the wire between the poles of a strong horseshoe magnet, subjecting 5.00 cm of it to a 1.75-T field, what force is exerted on this segment of wire?

Exercise:**Problem:**

(a) What is the angle between a wire carrying an 8.00-A current and the 1.20-T field it is in if 50.0 cm of the wire experiences a magnetic force of 2.40 N? (b) What is the force on the wire if it is rotated to make an angle of 90° with the field?

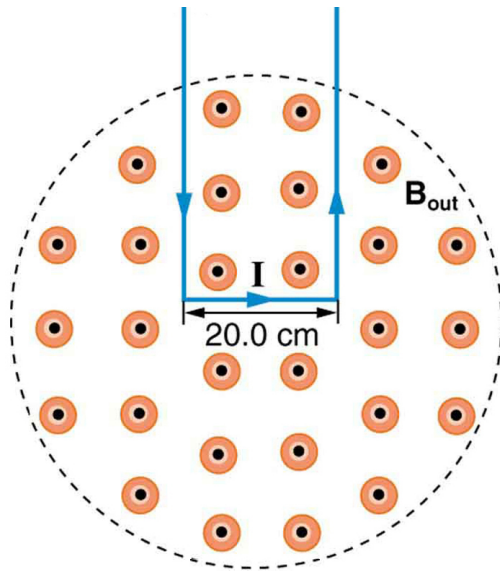
Solution:

(a) 30°

(b) 4.80 N

Exercise:**Problem:**

The force on the rectangular loop of wire in the magnetic field in [\[link\]](#) can be used to measure field strength. The field is uniform, and the plane of the loop is perpendicular to the field. (a) What is the direction of the magnetic force on the loop? Justify the claim that the forces on the sides of the loop are equal and opposite, independent of how much of the loop is in the field and do not affect the net force on the loop. (b) If a current of 5.00 A is used, what is the force per tesla on the 20.0-cm-wide loop?

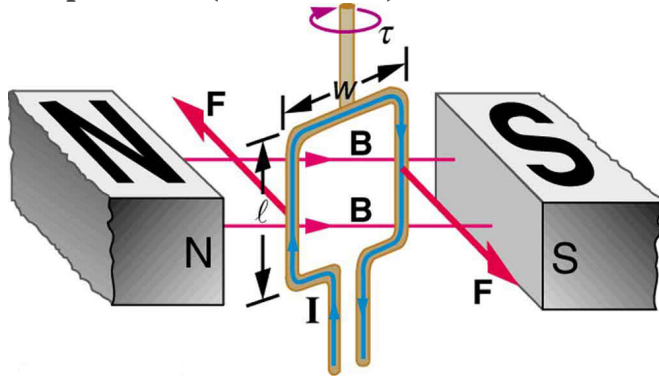


A rectangular loop of wire carrying a current is perpendicular to a magnetic field. The field is uniform in the region shown and is zero outside that region.

Torque on a Current Loop: Motors and Meters

- Describe how motors and meters work in terms of torque on a current loop.
- Calculate the torque on a current-carrying loop in a magnetic field.

Motors are the most common application of magnetic force on current-carrying wires. Motors have loops of wire in a magnetic field. When current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft. Electrical energy is converted to mechanical work in the process. (See [\[link\]](#).)



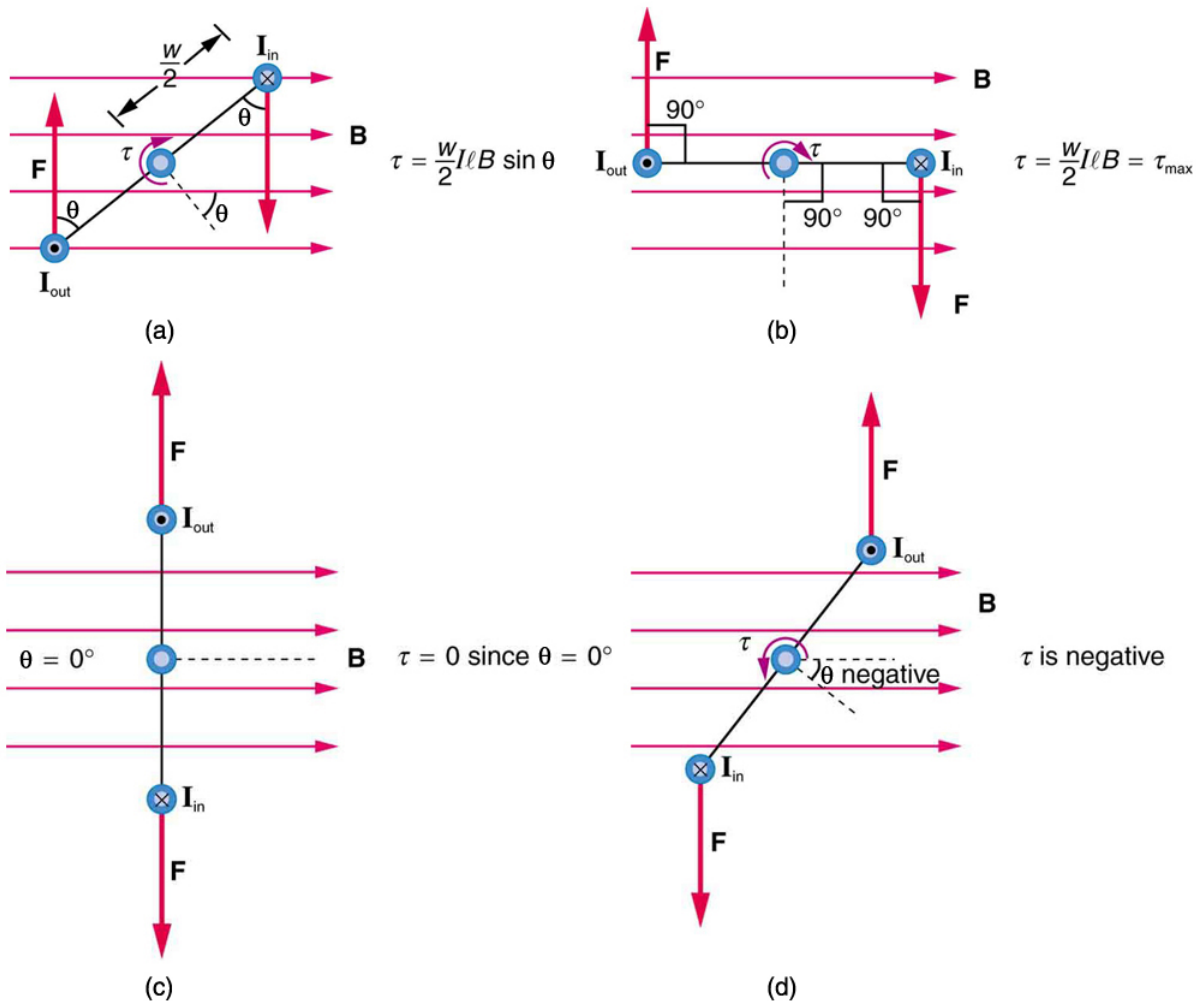
Torque on a current loop. A current-carrying loop of wire attached to a vertically rotating shaft feels magnetic forces that produce a clockwise torque as viewed from above.

Let us examine the force on each segment of the loop in [\[link\]](#) to find the torques produced about the axis of the vertical shaft. (This will lead to a useful equation for the torque on the loop.) We take the magnetic field to be uniform over the rectangular loop, which has width w and height l . First, we note that the forces on the top and bottom segments are vertical and, therefore, parallel to the shaft, producing no torque. Those vertical forces are equal in magnitude and opposite in direction, so that they also produce no net force on the loop. [\[link\]](#) shows views of the loop from above. Torque

is defined as $\tau = rF \sin \theta$, where F is the force, r is the distance from the pivot that the force is applied, and θ is the angle between r and F . As seen in [link](a), right hand rule 1 gives the forces on the sides to be equal in magnitude and opposite in direction, so that the net force is again zero. However, each force produces a clockwise torque. Since $r = w/2$, the torque on each vertical segment is $(w/2)F \sin \theta$, and the two add to give a total torque.

Equation:

$$\tau = \frac{w}{2} F \sin \theta + \frac{w}{2} F \sin \theta = wF \sin \theta$$



Top views of a current-carrying loop in a magnetic field. (a) The equation for torque is derived using this view. Note that the perpendicular to the loop makes an angle θ with the field that is the

same as the angle between $w/2$ and \mathbf{F} . (b) The maximum torque occurs when θ is a right angle and $\sin \theta = 1$. (c) Zero (minimum) torque occurs when θ is zero and $\sin \theta = 0$. (d) The torque reverses once the loop rotates past $\theta = 0$.

Now, each vertical segment has a length l that is perpendicular to B , so that the force on each is $F = IlB$. Entering F into the expression for torque yields

Equation:

$$\tau = wIlB \sin \theta.$$

If we have a multiple loop of N turns, we get N times the torque of one loop. Finally, note that the area of the loop is $A = wl$; the expression for the torque becomes

Equation:

$$\tau = NIAB \sin \theta.$$

This is the torque on a current-carrying loop in a uniform magnetic field. This equation can be shown to be valid for a loop of any shape. The loop carries a current I , has N turns, each of area A , and the perpendicular to the loop makes an angle θ with the field B . The net force on the loop is zero.

Example:

Calculating Torque on a Current-Carrying Loop in a Strong Magnetic Field

Find the maximum torque on a 100-turn square loop of a wire of 10.0 cm on a side that carries 15.0 A of current in a 2.00-T field.

Strategy

Torque on the loop can be found using $\tau = NIAB \sin \theta$. Maximum torque occurs when $\theta = 90^\circ$ and $\sin \theta = 1$.

Solution

For $\sin \theta = 1$, the maximum torque is

Equation:

$$\tau_{\max} = NIAB.$$

Entering known values yields

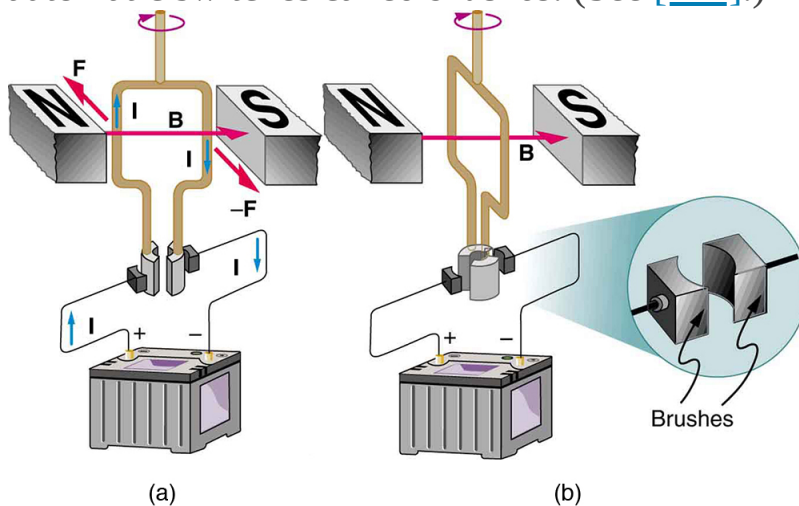
Equation:

$$\begin{aligned}\tau_{\max} &= (100)(15.0 \text{ A})(0.100 \text{ m}^2)(2.00 \text{ T}) \\ &= 30.0 \text{ N} \cdot \text{m}.\end{aligned}$$

Discussion

This torque is large enough to be useful in a motor.

The torque found in the preceding example is the maximum. As the coil rotates, the torque decreases to zero at $\theta = 0$. The torque then *reverses* its direction once the coil rotates past $\theta = 0$. (See [\[link\]](#)(d).) This means that, unless we do something, the coil will oscillate back and forth about equilibrium at $\theta = 0$. To get the coil to continue rotating in the same direction, we can reverse the current as it passes through $\theta = 0$ with automatic switches called *brushes*. (See [\[link\]](#).)

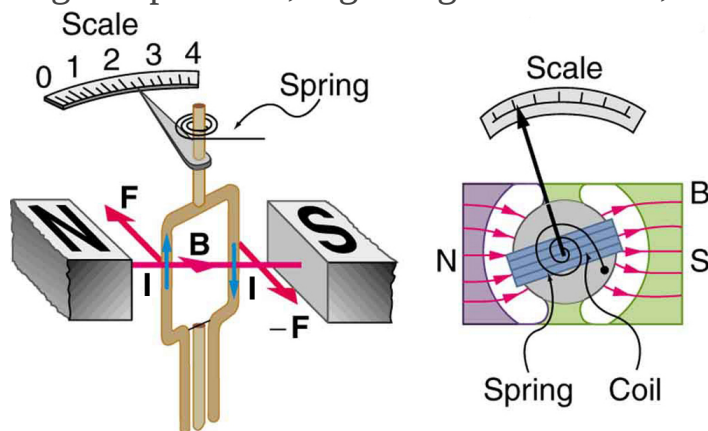


(a) As the angular momentum of the coil carries it through $\theta = 0$, the brushes reverse

the current to keep the torque clockwise. (b)

The coil will rotate continuously in the clockwise direction, with the current reversing each half revolution to maintain the clockwise torque.

Meters, such as those in analog fuel gauges on a car, are another common application of magnetic torque on a current-carrying loop. [\[link\]](#) shows that a meter is very similar in construction to a motor. The meter in the figure has its magnets shaped to limit the effect of θ by making B perpendicular to the loop over a large angular range. Thus the torque is proportional to I and not θ . A linear spring exerts a counter-torque that balances the current-produced torque. This makes the needle deflection proportional to I . If an exact proportionality cannot be achieved, the gauge reading can be calibrated. To produce a galvanometer for use in analog voltmeters and ammeters that have a low resistance and respond to small currents, we use a large loop area A , high magnetic field B , and low-resistance coils.



Meters are very similar to motors but only rotate through a part of a revolution. The magnetic poles of this meter are shaped to keep the component of B perpendicular to the loop constant, so that the torque does not depend on θ and the deflection

against the return spring is proportional only to the current I .

Section Summary

- The torque τ on a current-carrying loop of any shape in a uniform magnetic field. is

Equation:

$$\tau = NIAB \sin \theta,$$

where N is the number of turns, I is the current, A is the area of the loop, B is the magnetic field strength, and θ is the angle between the perpendicular to the loop and the magnetic field.

Conceptual Questions

Exercise:

Problem:

Draw a diagram and use RHR-1 to show that the forces on the top and bottom segments of the motor's current loop in [\[link\]](#) are vertical and produce no torque about the axis of rotation.

Problems & Exercises

Exercise:

Problem:

(a) By how many percent is the torque of a motor decreased if its permanent magnets lose 5.0% of their strength? (b) How many percent would the current need to be increased to return the torque to original values?

Solution:

(a) τ decreases by 5.00% if B decreases by 5.00%

(b) 5.26% increase

Exercise:

Problem:

(a) What is the maximum torque on a 150-turn square loop of wire 18.0 cm on a side that carries a 50.0-A current in a 1.60-T field? (b) What is the torque when θ is 10.9° ?

Exercise:

Problem:

Find the current through a loop needed to create a maximum torque of $9.00 \text{ N} \cdot \text{m}$. The loop has 50 square turns that are 15.0 cm on a side and is in a uniform 0.800-T magnetic field.

Solution:

10.0 A

Exercise:

Problem:

Calculate the magnetic field strength needed on a 200-turn square loop 20.0 cm on a side to create a maximum torque of $300 \text{ N} \cdot \text{m}$ if the loop is carrying 25.0 A.

Exercise:

Problem:

Since the equation for torque on a current-carrying loop is $\tau = NIAB \sin \theta$, the units of $\text{N} \cdot \text{m}$ must equal units of $\text{A} \cdot \text{m}^2 \text{ T}$. Verify this.

Solution:

$$A \cdot m^2 \cdot T = A \cdot m^2 \left(\frac{N}{A \cdot m} \right) = N \cdot m.$$

Exercise:

Problem:

(a) At what angle θ is the torque on a current loop 90.0% of maximum? (b) 50.0% of maximum? (c) 10.0% of maximum?

Exercise:

Problem:

A proton has a magnetic field due to its spin on its axis. The field is similar to that created by a circular current loop 0.650×10^{-15} m in radius with a current of 1.05×10^4 A (no kidding). Find the maximum torque on a proton in a 2.50-T field. (This is a significant torque on a small particle.)

Solution:

$$3.48 \times 10^{-26} \text{ N} \cdot \text{m}$$

Exercise:

Problem:

(a) A 200-turn circular loop of radius 50.0 cm is vertical, with its axis on an east-west line. A current of 100 A circulates clockwise in the loop when viewed from the east. The Earth's field here is due north, parallel to the ground, with a strength of 3.00×10^{-5} T. What are the direction and magnitude of the torque on the loop? (b) Does this device have any practical applications as a motor?

Exercise:

Problem:

Repeat [\[link\]](#), but with the loop lying flat on the ground with its current circulating counterclockwise (when viewed from above) in a location where the Earth's field is north, but at an angle 45.0° below the horizontal and with a strength of 6.00×10^{-5} T.

Solution:

(a) $0.666 \text{ N} \cdot \text{m}$ west

(b) This is not a very significant torque, so practical use would be limited. Also, the current would need to be alternated to make the loop rotate (otherwise it would oscillate).

Glossary

motor

loop of wire in a magnetic field; when current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft; electrical energy is converted to mechanical work in the process

meter

common application of magnetic torque on a current-carrying loop that is very similar in construction to a motor; by design, the torque is proportional to I and not θ , so the needle deflection is proportional to the current

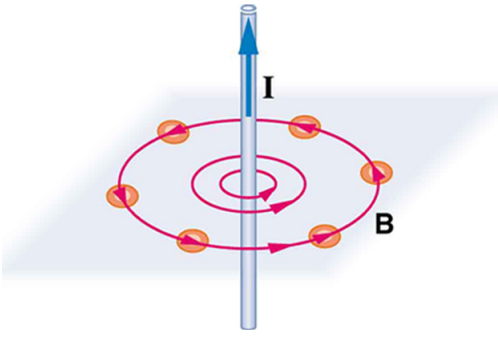
Magnetic Fields Produced by Currents: Ampere's Law

- Calculate current that produces a magnetic field.
- Use the right hand rule 2 to determine the direction of current or the direction of magnetic field loops.

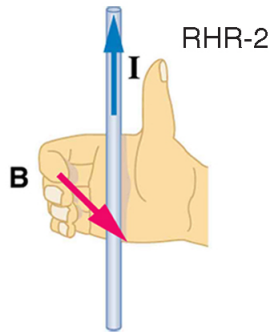
How much current is needed to produce a significant magnetic field, perhaps as strong as the Earth's field? Surveyors will tell you that overhead electric power lines create magnetic fields that interfere with their compass readings. Indeed, when Oersted discovered in 1820 that a current in a wire affected a compass needle, he was not dealing with extremely large currents. How does the shape of wires carrying current affect the shape of the magnetic field created? We noted earlier that a current loop created a magnetic field similar to that of a bar magnet, but what about a straight wire or a toroid (doughnut)? How is the direction of a current-created field related to the direction of the current? Answers to these questions are explored in this section, together with a brief discussion of the law governing the fields created by currents.

Magnetic Field Created by a Long Straight Current-Carrying Wire: Right Hand Rule 2

Magnetic fields have both direction and magnitude. As noted before, one way to explore the direction of a magnetic field is with compasses, as shown for a long straight current-carrying wire in [\[link\]](#). Hall probes can determine the magnitude of the field. The field around a long straight wire is found to be in circular loops. The **right hand rule 2** (RHR-2) emerges from this exploration and is valid for any current segment—*point the thumb in the direction of the current, and the fingers curl in the direction of the magnetic field loops* created by it.



(a)



(b)

(a) Compasses placed near a long straight current-carrying wire indicate that field lines form circular loops centered on the wire. (b) Right hand rule 2 states that, if the right hand thumb points in the direction of the current, the fingers curl in the direction of the field. This rule is consistent with the field mapped for the long straight wire and is valid for any current segment.

The **magnetic field strength (magnitude) produced by a long straight current-carrying wire** is found by experiment to be

Equation:

$$B = \frac{\mu_0 I}{2\pi r} \text{ (long straight wire),}$$

where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ is the **permeability of free space**. (μ_0 is one of the basic constants in nature. We will see later that μ_0 is related to the speed of light.) Since the wire is very long, the magnitude of the field depends only on distance from the wire r , not on position along the wire.

Example:

Calculating Current that Produces a Magnetic Field

Find the current in a long straight wire that would produce a magnetic field twice the strength of the Earth's at a distance of 5.0 cm from the wire.

Strategy

The Earth's field is about $5.0 \times 10^{-5} \text{ T}$, and so here B due to the wire is taken to be $1.0 \times 10^{-4} \text{ T}$. The equation $B = \frac{\mu_0 I}{2\pi r}$ can be used to find I , since all other quantities are known.

Solution

Solving for I and entering known values gives

Equation:

$$\begin{aligned} I &= \frac{2\pi r B}{\mu_0} = \frac{2\pi(5.0 \times 10^{-2} \text{ m})(1.0 \times 10^{-4} \text{ T})}{4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}} \\ &= 25 \text{ A.} \end{aligned}$$

Discussion

So a moderately large current produces a significant magnetic field at a distance of 5.0 cm from a long straight wire. Note that the answer is stated to only two digits, since the Earth's field is specified to only two digits in this example.

Ampere's Law and Others

The magnetic field of a long straight wire has more implications than you might at first suspect. *Each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment.* The formal statement of the direction and magnitude of the field due to each segment is called the **Biot-Savart law**. Integral calculus is needed to sum the field for an arbitrary shape current. This results in a more complete law, called **Ampere's law**, which relates magnetic field and current in a general way. Ampere's law in turn is a part of **Maxwell's equations**, which give a complete theory of all electromagnetic phenomena. Considerations of how Maxwell's equations appear to different observers led to the modern theory of relativity, and the realization that electric and magnetic fields are different manifestations of the same thing. Most of this is beyond the scope of this text in both mathematical level, requiring calculus, and in the amount of space that can be devoted to it. But for the interested student, and particularly for those who continue in physics, engineering, or similar pursuits, delving into these matters further will reveal descriptions of nature that are elegant as well as profound. In this text, we shall keep the general features in mind, such as RHR-2 and the rules for magnetic field lines listed in [Magnetic Fields and Magnetic Field Lines](#), while concentrating on the fields created in certain important situations.

Note:

Making Connections: Relativity

Hearing all we do about Einstein, we sometimes get the impression that he invented relativity out of nothing. On the contrary, one of Einstein's motivations was to solve difficulties in knowing how different observers see magnetic and electric fields.

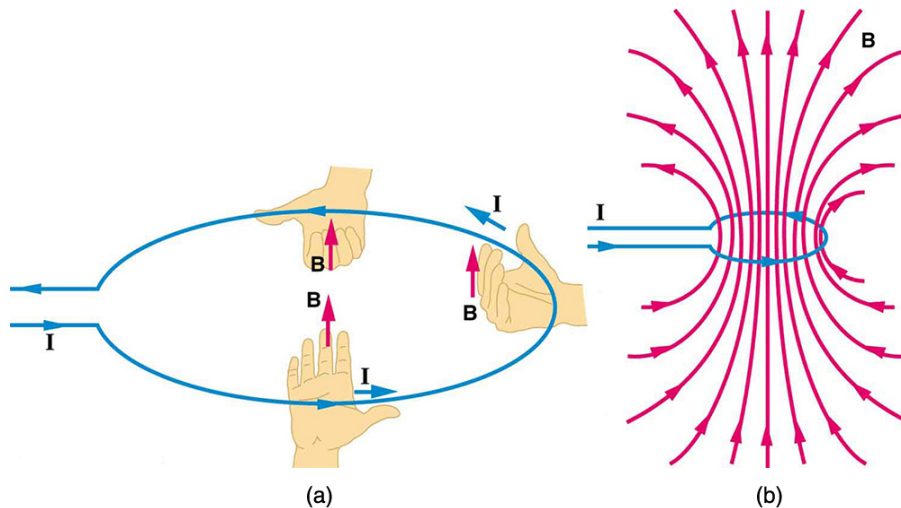
Magnetic Field Produced by a Current-Carrying Circular Loop

The magnetic field near a current-carrying loop of wire is shown in [\[link\]](#). Both the direction and the magnitude of the magnetic field produced by a current-carrying loop are complex. RHR-2 can be used to give the direction of the field near the loop, but mapping with compasses and the rules about field lines given in [Magnetic Fields and Magnetic Field Lines](#) are needed for more detail. There is a simple formula for the **magnetic field strength at the center of a circular loop**. It is

Equation:

$$B = \frac{\mu_0 I}{2R} \text{ (at center of loop),}$$

where R is the radius of the loop. This equation is very similar to that for a straight wire, but it is valid *only* at the center of a circular loop of wire. The similarity of the equations does indicate that similar field strength can be obtained at the center of a loop. One way to get a larger field is to have N loops; then, the field is $B = N\mu_0 I / (2R)$. Note that the larger the loop, the smaller the field at its center, because the current is farther away.

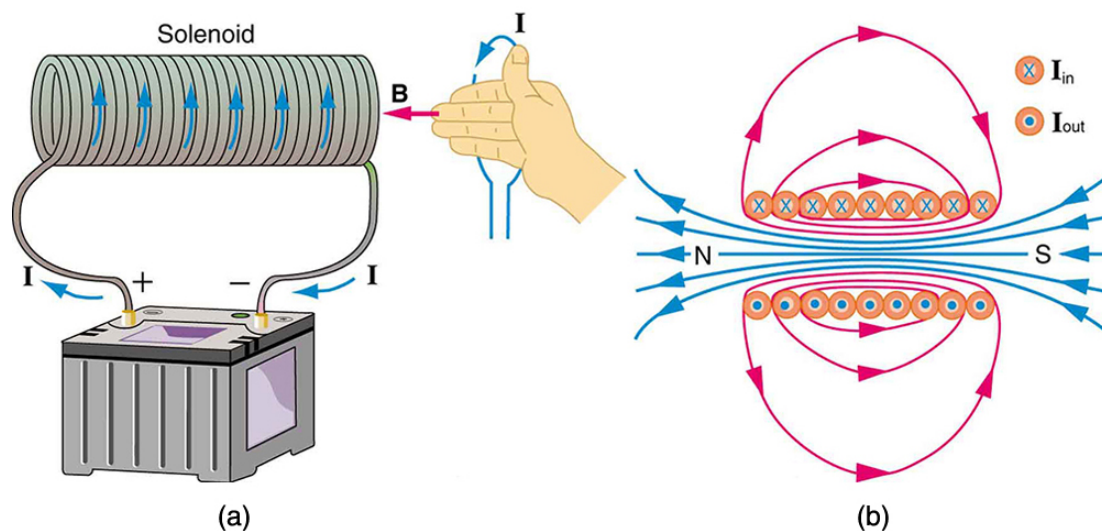


- (a) RHR-2 gives the direction of the magnetic field inside and outside a current-carrying loop. (b) More detailed mapping with compasses or with a

Hall probe completes the picture. The field is similar to that of a bar magnet.

Magnetic Field Produced by a Current-Carrying Solenoid

A **solenoid** is a long coil of wire (with many turns or loops, as opposed to a flat loop). Because of its shape, the field inside a solenoid can be very uniform, and also very strong. The field just outside the coils is nearly zero. [\[link\]](#) shows how the field looks and how its direction is given by RHR-2.



(a) Because of its shape, the field inside a solenoid of length l is remarkably uniform in magnitude and direction, as indicated by the straight and uniformly spaced field lines. The field outside the coils is nearly zero. (b) This cutaway shows the magnetic field generated by the current in the solenoid.

The magnetic field inside of a current-carrying solenoid is very uniform in direction and magnitude. Only near the ends does it begin to weaken and change direction. The field outside has similar complexities to flat loops

and bar magnets, but the **magnetic field strength inside a solenoid** is simply

Equation:

$$B = \mu_0 n I \quad (\text{inside a solenoid}),$$

where n is the number of loops per unit length of the solenoid ($n = N/l$, with N being the number of loops and l the length). Note that B is the field strength anywhere in the uniform region of the interior and not just at the center. Large uniform fields spread over a large volume are possible with solenoids, as [\[link\]](#) implies.

Example:

Calculating Field Strength inside a Solenoid

What is the field inside a 2.00-m-long solenoid that has 2000 loops and carries a 1600-A current?

Strategy

To find the field strength inside a solenoid, we use $B = \mu_0 n I$. First, we note the number of loops per unit length is

Equation:

$$n = \frac{N}{l} = \frac{2000}{2.00 \text{ m}} = 1000 \text{ m}^{-1} = 10 \text{ cm}^{-1}.$$

Solution

Substituting known values gives

Equation:

$$\begin{aligned} B &= \mu_0 n I = (4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}) (1000 \text{ m}^{-1}) (1600 \text{ A}) \\ &= 2.01 \text{ T}. \end{aligned}$$

Discussion

This is a large field strength that could be established over a large-diameter solenoid, such as in medical uses of magnetic resonance imaging (MRI).

The very large current is an indication that the fields of this strength are not

easily achieved, however. Such a large current through 1000 loops squeezed into a meter's length would produce significant heating. Higher currents can be achieved by using superconducting wires, although this is expensive. There is an upper limit to the current, since the superconducting state is disrupted by very large magnetic fields.

There are interesting variations of the flat coil and solenoid. For example, the toroidal coil used to confine the reactive particles in tokamaks is much like a solenoid bent into a circle. The field inside a toroid is very strong but circular. Charged particles travel in circles, following the field lines, and collide with one another, perhaps inducing fusion. But the charged particles do not cross field lines and escape the toroid. A whole range of coil shapes are used to produce all sorts of magnetic field shapes. Adding ferromagnetic materials produces greater field strengths and can have a significant effect on the shape of the field. Ferromagnetic materials tend to trap magnetic fields (the field lines bend into the ferromagnetic material, leaving weaker fields outside it) and are used as shields for devices that are adversely affected by magnetic fields, including the Earth's magnetic field.

Note:

PhET Explorations: Generator

Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

[Generato](#)

[r](#)

Section Summary

- The strength of the magnetic field created by current in a long straight wire is given by

Equation:

$$B = \frac{\mu_0 I}{2\pi r} (\text{long straight wire}),$$

where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ is the permeability of free space.

- The direction of the magnetic field created by a long straight wire is given by right hand rule 2 (RHR-2): *Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops* created by it.
- The magnetic field created by current following any path is the sum (or integral) of the fields due to segments along the path (magnitude and direction as for a straight wire), resulting in a general relationship between current and field known as Ampere's law.
- The magnetic field strength at the center of a circular loop is given by

Equation:

$$B = \frac{\mu_0 I}{2R} (\text{at center of loop}),$$

where R is the radius of the loop. This equation becomes $B = \mu_0 nI / (2R)$ for a flat coil of N loops. RHR-2 gives the direction of the field about the loop. A long coil is called a solenoid.

- The magnetic field strength inside a solenoid is

Equation:

$$B = \mu_0 nI \text{ (inside a solenoid),}$$

where n is the number of loops per unit length of the solenoid. The field inside is very uniform in magnitude and direction.

Conceptual Questions

Exercise:

Problem:

Make a drawing and use RHR-2 to find the direction of the magnetic field of a current loop in a motor (such as in [\[link\]](#)). Then show that the direction of the torque on the loop is the same as produced by like poles repelling and unlike poles attracting.

Glossary

right hand rule 2 (RHR-2)

a rule to determine the direction of the magnetic field induced by a current-carrying wire: Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops

magnetic field strength (magnitude) produced by a long straight current-carrying wire

defined as $B = \frac{\mu_0 I}{2\pi r}$, where I is the current, r is the shortest distance to the wire, and μ_0 is the permeability of free space

permeability of free space

the measure of the ability of a material, in this case free space, to support a magnetic field; the constant $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$

magnetic field strength at the center of a circular loop

defined as $B = \frac{\mu_0 I}{2R}$ where R is the radius of the loop

solenoid

a thin wire wound into a coil that produces a magnetic field when an electric current is passed through it

magnetic field strength inside a solenoid

defined as $B = \mu_0 n I$ where n is the number of loops per unit length of the solenoid ($n = N/l$, with N being the number of loops and l the length)

Biot-Savart law

a physical law that describes the magnetic field generated by an electric current in terms of a specific equation

Ampere's law

the physical law that states that the magnetic field around an electric current is proportional to the current; each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment

Maxwell's equations

a set of four equations that describe electromagnetic phenomena

Magnetic Force between Two Parallel Conductors

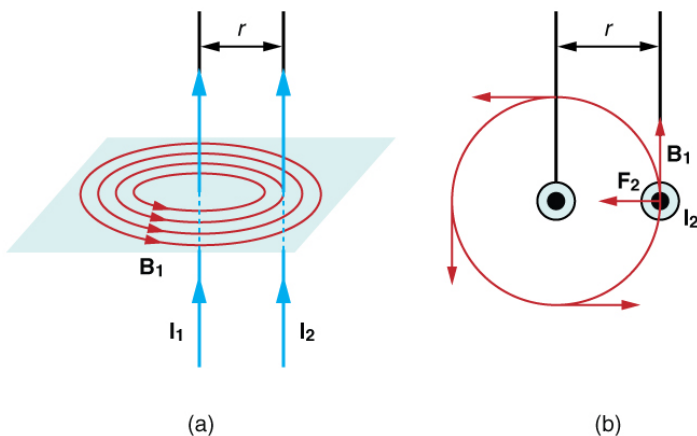
- Describe the effects of the magnetic force between two conductors.
- Calculate the force between two parallel conductors.

You might expect that there are significant forces between current-carrying wires, since ordinary currents produce significant magnetic fields and these fields exert significant forces on ordinary currents. But you might not expect that the force between wires is used to *define* the ampere. It might also surprise you to learn that this force has something to do with why large circuit breakers burn up when they attempt to interrupt large currents.

The force between two long straight and parallel conductors separated by a distance r can be found by applying what we have developed in preceding sections. [\[link\]](#) shows the wires, their currents, the fields they create, and the subsequent forces they exert on one another. Let us consider the field produced by wire 1 and the force it exerts on wire 2 (call the force F_2). The field due to I_1 at a distance r is given to be

Equation:

$$B_1 = \frac{\mu_0 I_1}{2\pi r}.$$



(a) The magnetic field produced by a long straight conductor is perpendicular to a parallel conductor, as indicated by RHR-2. (b) A view

from above of the two wires shown in (a), with one magnetic field line shown for each wire. RHR-1 shows that the force between the parallel conductors is attractive when the currents are in the same direction. A similar analysis shows that the force is repulsive between currents in opposite directions.

This field is uniform along wire 2 and perpendicular to it, and so the force F_2 it exerts on wire 2 is given by $F = IlB \sin \theta$ with $\sin \theta = 1$:

Equation:

$$F_2 = I_2 l B_1.$$

By Newton's third law, the forces on the wires are equal in magnitude, and so we just write F for the magnitude of F_2 . (Note that $F_1 = -F_2$.) Since the wires are very long, it is convenient to think in terms of F/l , the force per unit length. Substituting the expression for B_1 into the last equation and rearranging terms gives

Equation:

$$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$

F/l is the force per unit length between two parallel currents I_1 and I_2 separated by a distance r . The force is attractive if the currents are in the same direction and repulsive if they are in opposite directions.

This force is responsible for the *pinch effect* in electric arcs and plasmas. The force exists whether the currents are in wires or not. In an electric arc, where currents are moving parallel to one another, there is an attraction that squeezes currents into a smaller tube. In large circuit breakers, like those

used in neighborhood power distribution systems, the pinch effect can concentrate an arc between plates of a switch trying to break a large current, burn holes, and even ignite the equipment. Another example of the pinch effect is found in the solar plasma, where jets of ionized material, such as solar flares, are shaped by magnetic forces.

The *operational definition of the ampere* is based on the force between current-carrying wires. Note that for parallel wires separated by 1 meter with each carrying 1 ampere, the force per meter is

Equation:

$$\frac{F}{l} = \frac{(4\pi \times 10^{-7} \text{ T} \cdot \text{m/A})(1 \text{ A})^2}{(2\pi)(1 \text{ m})} = 2 \times 10^{-7} \text{ N/m}.$$

Since μ_0 is exactly $4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ by definition, and because $1 \text{ T} = 1 \text{ N}/(\text{A} \cdot \text{m})$, the force per meter is exactly $2 \times 10^{-7} \text{ N/m}$. This is the basis of the operational definition of the ampere.

Note:

The Ampere

The official definition of the ampere is:

One ampere of current through each of two parallel conductors of infinite length, separated by one meter in empty space free of other magnetic fields, causes a force of exactly $2 \times 10^{-7} \text{ N/m}$ on each conductor.

Infinite-length straight wires are impractical and so, in practice, a current balance is constructed with coils of wire separated by a few centimeters. Force is measured to determine current. This also provides us with a method for measuring the coulomb. We measure the charge that flows for a current of one ampere in one second. That is, $1 \text{ C} = 1 \text{ A} \cdot \text{s}$. For both the ampere and the coulomb, the method of measuring force between conductors is the most accurate in practice.

Section Summary

- The force between two parallel currents I_1 and I_2 , separated by a distance r , has a magnitude per unit length given by

Equation:

$$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$

- The force is attractive if the currents are in the same direction, repulsive if they are in opposite directions.

Conceptual Questions

Exercise:

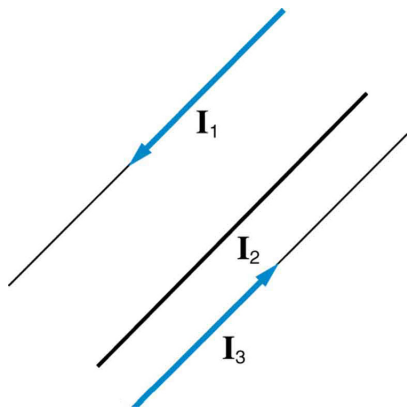
Problem:

Is the force attractive or repulsive between the hot and neutral lines hung from power poles? Why?

Exercise:

Problem:

If you have three parallel wires in the same plane, as in [\[link\]](#), with currents in the outer two running in opposite directions, is it possible for the middle wire to be repelled by both? Attracted by both? Explain.



Three parallel

coplanar wires with
currents in the
outer two in
opposite directions.

Exercise:

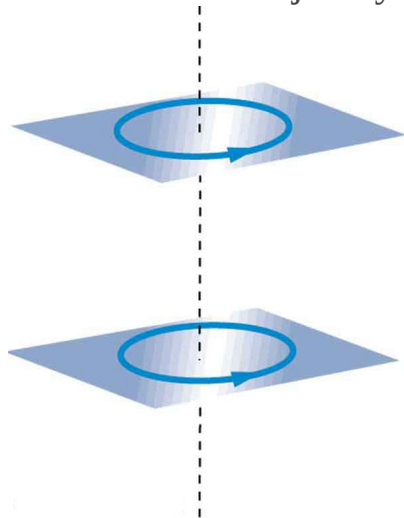
Problem:

Suppose two long straight wires run perpendicular to one another without touching. Does one exert a net force on the other? If so, what is its direction? Does one exert a net torque on the other? If so, what is its direction? Justify your responses by using the right hand rules.

Exercise:

Problem:

Use the right hand rules to show that the force between the two loops in [\[link\]](#) is attractive if the currents are in the same direction and repulsive if they are in opposite directions. Is this consistent with like poles of the loops repelling and unlike poles of the loops attracting? Draw sketches to justify your answers.



Two loops of wire
carrying currents

can exert forces
and torques on one
another.

Exercise:

Problem:

If one of the loops in [\[link\]](#) is tilted slightly relative to the other and their currents are in the same direction, what are the directions of the torques they exert on each other? Does this imply that the poles of the bar magnet-like fields they create will line up with each other if the loops are allowed to rotate?

Exercise:

Problem:

Electric field lines can be shielded by the Faraday cage effect. Can we have magnetic shielding? Can we have gravitational shielding?

Problems & Exercises

Exercise:

Problem:

- (a) The hot and neutral wires supplying DC power to a light-rail commuter train carry 800 A and are separated by 75.0 cm. What is the magnitude and direction of the force between 50.0 m of these wires?
- (b) Discuss the practical consequences of this force, if any.

Solution:

- (a) 8.53 N, repulsive
- (b) This force is repulsive and therefore there is never a risk that the two wires will touch and short circuit.

Exercise:**Problem:**

The force per meter between the two wires of a jumper cable being used to start a stalled car is 0.225 N/m. (a) What is the current in the wires, given they are separated by 2.00 cm? (b) Is the force attractive or repulsive?

Exercise:**Problem:**

A 2.50-m segment of wire supplying current to the motor of a submerged submarine carries 1000 A and feels a 4.00-N repulsive force from a parallel wire 5.00 cm away. What is the direction and magnitude of the current in the other wire?

Solution:

400 A in the opposite direction

Exercise:**Problem:**

The wire carrying 400 A to the motor of a commuter train feels an attractive force of 4.00×10^{-3} N/m due to a parallel wire carrying 5.00 A to a headlight. (a) How far apart are the wires? (b) Are the currents in the same direction?

Exercise:**Problem:**

An AC appliance cord has its hot and neutral wires separated by 3.00 mm and carries a 5.00-A current. (a) What is the average force per meter between the wires in the cord? (b) What is the maximum force per meter between the wires? (c) Are the forces attractive or repulsive? (d) Do appliance cords need any special design features to compensate for these forces?

Solution:

(a) $1.67 \times 10^{-3} \text{ N/m}$

(b) $3.33 \times 10^{-3} \text{ N/m}$

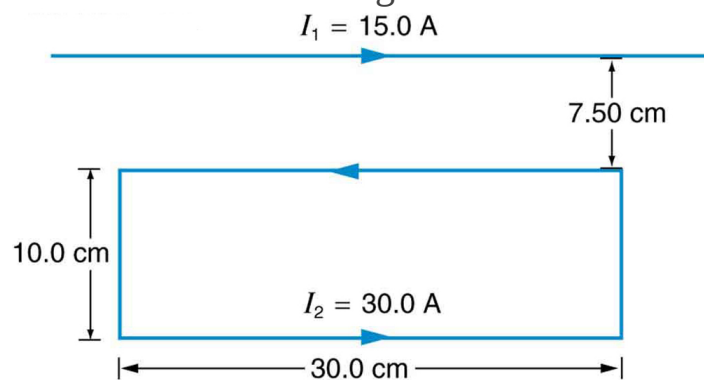
(c) Repulsive

(d) No, these are very small forces

Exercise:

Problem:

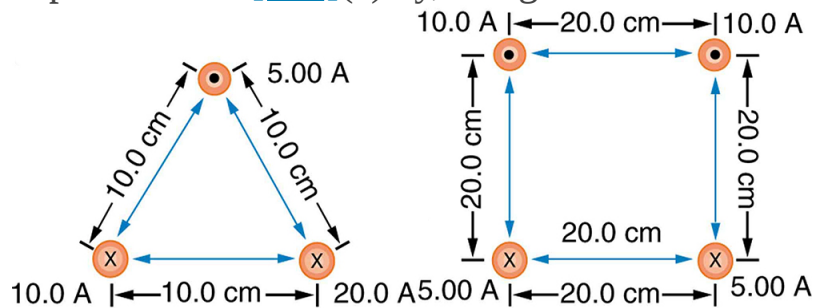
[\[link\]](#) shows a long straight wire near a rectangular current loop. What is the direction and magnitude of the total force on the loop?



Exercise:

Problem:

Find the direction and magnitude of the force that each wire experiences in [\[link\]](#)(a) by, using vector addition.



Solution:

(a) Top wire: $2.65 \times 10^{-4} \text{ N/m}$, 10.9° to left of up

(b) Lower left wire: $3.61 \times 10^{-4} \text{ N/m}$, 13.9° down from right

(c) Lower right wire: $3.46 \times 10^{-4} \text{ N/m}$, 30.0° down from left

Exercise:**Problem:**

Find the direction and magnitude of the force that each wire experiences in [\[link\]](#)(b), using vector addition.

More Applications of Magnetism

- Describe some applications of magnetism.

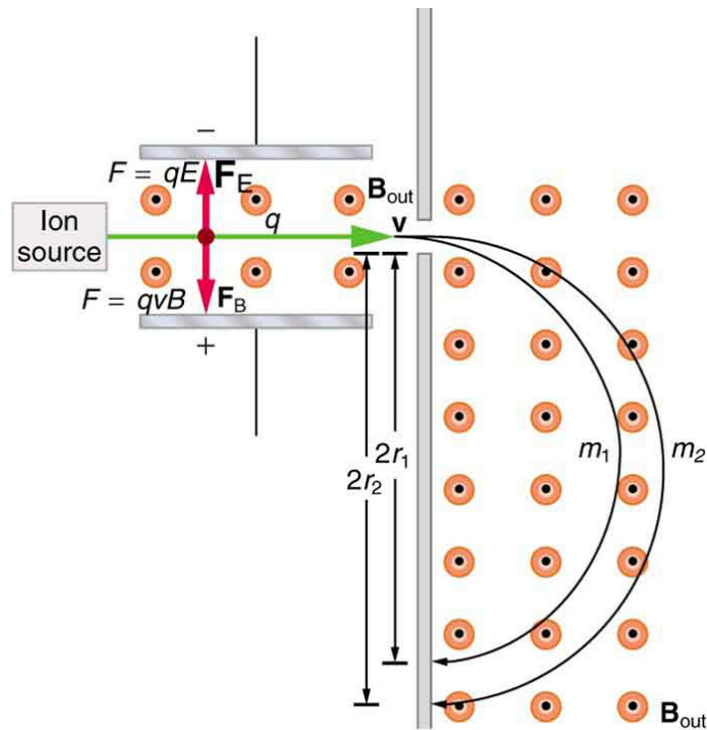
Mass Spectrometry

The curved paths followed by charged particles in magnetic fields can be put to use. A charged particle moving perpendicular to a magnetic field travels in a circular path having a radius r .

Equation:

$$r = \frac{mv}{qB}$$

It was noted that this relationship could be used to measure the mass of charged particles such as ions. A mass spectrometer is a device that measures such masses. Most mass spectrometers use magnetic fields for this purpose, although some of them have extremely sophisticated designs. Since there are five variables in the relationship, there are many possibilities. However, if v , q , and B can be fixed, then the radius of the path r is simply proportional to the mass m of the charged particle. Let us examine one such mass spectrometer that has a relatively simple design. (See [\[link\]](#).) The process begins with an ion source, a device like an electron gun. The ion source gives ions their charge, accelerates them to some velocity v , and directs a beam of them into the next stage of the spectrometer. This next region is a *velocity selector* that only allows particles with a particular value of v to get through.



This mass spectrometer uses a velocity selector to fix v so that the radius of the path is proportional to mass.

The velocity selector has both an electric field and a magnetic field, perpendicular to one another, producing forces in opposite directions on the ions. Only those ions for which the forces balance travel in a straight line into the next region. If the forces balance, then the electric force $F = qE$ equals the magnetic force $F = qvB$, so that $qE = qvB$. Noting that q

Equation:

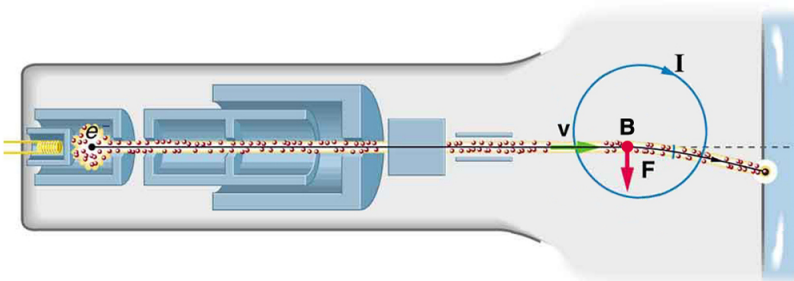
$$v = \frac{E}{B}$$

is the velocity particles must have to make it through the velocity selector, and further, that v can be selected by varying E and B . In the final region, there is only a uniform magnetic field, and so the charged particles move in circular arcs with radii proportional to particle mass. The paths also depend on charge q , but since q is in multiples of electron charges, it is easy to determine and to discriminate between ions in different charge states.

Mass spectrometry today is used extensively in chemistry and biology laboratories to identify chemical and biological substances according to their mass-to-charge ratios. In medicine, mass spectrometers are used to measure the concentration of isotopes used as tracers. Usually, biological molecules such as proteins are very large, so they are broken down into smaller fragments before analyzing. Recently, large virus particles have been analyzed as a whole on mass spectrometers. Sometimes a gas chromatograph or high-performance liquid chromatograph provides an initial separation of the large molecules, which are then input into the mass spectrometer.

Cathode Ray Tubes—CRTs—and the Like

What do non-flat-screen TVs, old computer monitors, x-ray machines, and the 2-mile-long Stanford Linear Accelerator have in common? All of them accelerate electrons, making them different versions of the electron gun. Many of these devices use magnetic fields to steer the accelerated electrons. [\[link\]](#) shows the construction of the type of cathode ray tube (CRT) found in some TVs, oscilloscopes, and old computer monitors. Two pairs of coils are used to steer the electrons, one vertically and the other horizontally, to their desired destination.



The cathode ray tube (CRT) is so named because rays of electrons originate at the cathode in the electron gun. Magnetic coils are used to steer the beam in many CRTs. In this case, the beam is moved down. Another pair of horizontal coils would steer the beam horizontally.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is one of the most useful and rapidly growing medical imaging tools. It non-invasively produces two-dimensional and three-dimensional images of the body that provide important medical information with none of the hazards of x-rays. MRI is based on an effect called **nuclear magnetic resonance (NMR)** in which an externally applied magnetic field interacts with the nuclei of certain atoms, particularly those of hydrogen (protons). These nuclei possess their own small magnetic fields, similar to those of electrons and the current loops discussed earlier in this chapter.

When placed in an external magnetic field, such nuclei experience a torque that pushes or aligns the nuclei into one of two new energy states—depending on the orientation of its spin (analogous to the N pole and S pole in a bar magnet). Transitions from the lower to higher energy state can be achieved by using an external radio frequency signal to “flip” the orientation of the small magnets. (This is actually a quantum mechanical process. The direction of the nuclear magnetic field is quantized as is energy in the radio waves. We will return to these topics in later chapters.) The specific frequency of the radio waves that are absorbed and reemitted depends sensitively on the type of nucleus, the chemical environment, and the external magnetic field strength. Therefore, this is a *resonance* phenomenon in which *nuclei* in a *magnetic* field act like resonators (analogous to those discussed in the treatment of sound in [Oscillatory](#)

[Motion and Waves](#)) that absorb and reemit only certain frequencies. Hence, the phenomenon is named *nuclear magnetic resonance (NMR)*.

NMR has been used for more than 50 years as an analytical tool. It was formulated in 1946 by F. Bloch and E. Purcell, with the 1952 Nobel Prize in Physics going to them for their work. Over the past two decades, NMR has been developed to produce detailed images in a process now called magnetic resonance imaging (MRI), a name coined to avoid the use of the word “nuclear” and the concomitant implication that nuclear radiation is involved. (It is not.) The 2003 Nobel Prize in Medicine went to P. Lauterbur and P. Mansfield for their work with MRI applications.

The largest part of the MRI unit is a superconducting magnet that creates a magnetic field, typically between 1 and 2 T in strength, over a relatively large volume. MRI images can be both highly detailed and informative about structures and organ functions. It is helpful that normal and non-normal tissues respond differently for slight changes in the magnetic field. In most medical images, the protons that are hydrogen nuclei are imaged. (About 2/3 of the atoms in the body are hydrogen.) Their location and density give a variety of medically useful information, such as organ function, the condition of tissue (as in the brain), and the shape of structures, such as vertebral disks and knee-joint surfaces. MRI can also be used to follow the movement of certain ions across membranes, yielding information on active transport, osmosis, dialysis, and other phenomena. With excellent spatial resolution, MRI can provide information about tumors, strokes, shoulder injuries, infections, etc.

An image requires position information as well as the density of a nuclear type (usually protons). By varying the magnetic field slightly over the volume to be imaged, the resonant frequency of the protons is made to vary with position. Broadcast radio frequencies are swept over an appropriate range and nuclei absorb and reemit them only if the nuclei are in a magnetic field with the correct strength. The imaging receiver gathers information through the body almost point by point, building up a tissue map. The reception of reemitted radio waves as a function of frequency thus gives position information. These “slices” or cross sections through the body are only several mm thick. The intensity of the reemitted radio waves is

proportional to the concentration of the nuclear type being flipped, as well as information on the chemical environment in that area of the body.

Various techniques are available for enhancing contrast in images and for obtaining more information. Scans called T1, T2, or proton density scans rely on different relaxation mechanisms of nuclei. Relaxation refers to the time it takes for the protons to return to equilibrium after the external field is turned off. This time depends upon tissue type and status (such as inflammation).

While MRI images are superior to x rays for certain types of tissue and have none of the hazards of x rays, they do not completely supplant x-ray images. MRI is less effective than x rays for detecting breaks in bone, for example, and in imaging breast tissue, so the two diagnostic tools complement each other. MRI images are also expensive compared to simple x-ray images and tend to be used most often where they supply information not readily obtained from x rays. Another disadvantage of MRI is that the patient is totally enclosed with detectors close to the body for about 30 minutes or more, leading to claustrophobia. It is also difficult for the obese patient to be in the magnet tunnel. New “open-MRI” machines are now available in which the magnet does not completely surround the patient.

Over the last decade, the development of much faster scans, called “functional MRI” (fMRI), has allowed us to map the functioning of various regions in the brain responsible for thought and motor control. This technique measures the change in blood flow for activities (thought, experiences, action) in the brain. The nerve cells increase their consumption of oxygen when active. Blood hemoglobin releases oxygen to active nerve cells and has somewhat different magnetic properties when oxygenated than when deoxygenated. With MRI, we can measure this and detect a blood oxygen-dependent signal. Most of the brain scans today use fMRI.

Other Medical Uses of Magnetic Fields

Currents in nerve cells and the heart create magnetic fields like any other currents. These can be measured but with some difficulty since their strengths are about 10^{-6} to 10^{-8} less than the Earth’s magnetic field. Recording of the heart’s magnetic field as it beats is called a

magnetocardiogram (MCG), while measurements of the brain's magnetic field is called a **magnetoencephalogram (MEG)**. Both give information that differs from that obtained by measuring the electric fields of these organs (ECGs and EEGs), but they are not yet of sufficient importance to make these difficult measurements common.

In both of these techniques, the sensors do not touch the body. MCG can be used in fetal studies, and is probably more sensitive than echocardiography. MCG also looks at the heart's electrical activity whose voltage output is too small to be recorded by surface electrodes as in EKG. It has the potential of being a rapid scan for early diagnosis of cardiac ischemia (obstruction of blood flow to the heart) or problems with the fetus.

MEG can be used to identify abnormal electrical discharges in the brain that produce weak magnetic signals. Therefore, it looks at brain activity, not just brain structure. It has been used for studies of Alzheimer's disease and epilepsy. Advances in instrumentation to measure very small magnetic fields have allowed these two techniques to be used more in recent years. What is used is a sensor called a SQUID, for superconducting quantum interference device. This operates at liquid helium temperatures and can measure magnetic fields thousands of times smaller than the Earth's.

Finally, there is a burgeoning market for magnetic cures in which magnets are applied in a variety of ways to the body, from magnetic bracelets to magnetic mattresses. The best that can be said for such practices is that they are apparently harmless, unless the magnets get close to the patient's computer or magnetic storage disks. Claims are made for a broad spectrum of benefits from cleansing the blood to giving the patient more energy, but clinical studies have not verified these claims, nor is there an identifiable mechanism by which such benefits might occur.

Note:

PhET Explorations: Magnet and Compass

Ever wonder how a compass worked to point you to the Arctic? Explore the interactions between a compass and bar magnet, and then add the Earth and find the surprising answer! Vary the magnet's strength, and see how

things change both inside and outside. Use the field meter to measure how the magnetic field changes.

<https://archive.cnx.org/specials/5ca3e2cc-ae74-11e5-b6d3-f3c228f04b5c/magnet-and-compass/#sim-bar-magnet>

Section Summary

- Crossed (perpendicular) electric and magnetic fields act as a velocity filter, giving equal and opposite forces on any charge with velocity perpendicular to the fields and of magnitude

Equation:

$$v = \frac{E}{B}.$$

Conceptual Questions

Exercise:

Problem:

Measurements of the weak and fluctuating magnetic fields associated with brain activity are called magnetoencephalograms (MEGs). Do the brain's magnetic fields imply coordinated or uncoordinated nerve impulses? Explain.

Exercise:

Problem:

Discuss the possibility that a Hall voltage would be generated on the moving heart of a patient during MRI imaging. Also discuss the same effect on the wires of a pacemaker. (The fact that patients with pacemakers are not given MRIs is significant.)

Exercise:

Problem:

A patient in an MRI unit turns his head quickly to one side and experiences momentary dizziness and a strange taste in his mouth. Discuss the possible causes.

Exercise:**Problem:**

You are told that in a certain region there is either a uniform electric or magnetic field. What measurement or observation could you make to determine the type? (Ignore the Earth's magnetic field.)

Exercise:**Problem:**

An example of magnetohydrodynamics (MHD) comes from the flow of a river (salty water). This fluid interacts with the Earth's magnetic field to produce a potential difference between the two river banks. How would you go about calculating the potential difference?

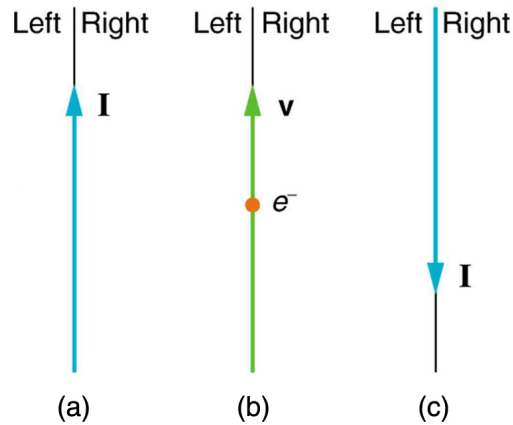
Exercise:**Problem:**

Draw gravitational field lines between 2 masses, electric field lines between a positive and a negative charge, electric field lines between 2 positive charges and magnetic field lines around a magnet. Qualitatively describe the differences between the fields and the entities responsible for the field lines.

Problems & Exercises**Exercise:**

Problem:

Indicate whether the magnetic field created in each of the three situations shown in [\[link\]](#) is into or out of the page on the left and right of the current.

**Solution:**

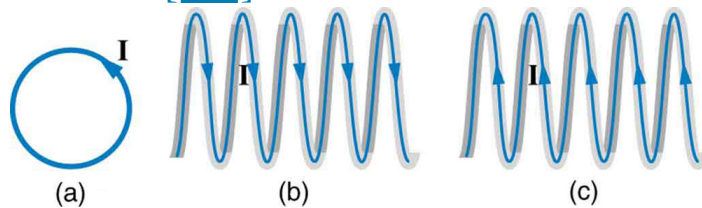
(a) right-into page, left-out of page

(b) right-out of page, left-into page

(c) right-out of page, left-into page

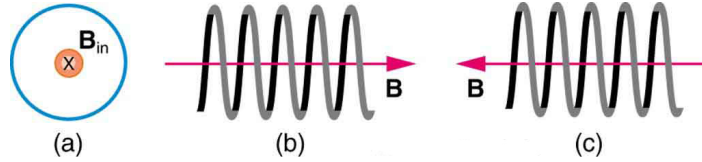
Exercise:**Problem:**

What are the directions of the fields in the center of the loop and coils shown in [\[link\]](#)?

**Exercise:**

Problem:

What are the directions of the currents in the loop and coils shown in [\[link\]](#)?

**Solution:**

- (a) clockwise
- (b) clockwise as seen from the left
- (c) clockwise as seen from the right

Exercise:**Problem:**

To see why an MRI utilizes iron to increase the magnetic field created by a coil, calculate the current needed in a 400-loop-per-meter circular coil 0.660 m in radius to create a 1.20-T field (typical of an MRI instrument) at its center with no iron present. The magnetic field of a proton is approximately like that of a circular current loop 0.650×10^{-15} m in radius carrying 1.05×10^4 A. What is the field at the center of such a loop?

Solution:

$$1.01 \times 10^{13} \text{ T}$$

Exercise:**Problem:**

Inside a motor, 30.0 A passes through a 250-turn circular loop that is 10.0 cm in radius. What is the magnetic field strength created at its center?

Exercise:**Problem:**

Nonnuclear submarines use batteries for power when submerged. (a) Find the magnetic field 50.0 cm from a straight wire carrying 1200 A from the batteries to the drive mechanism of a submarine. (b) What is the field if the wires to and from the drive mechanism are side by side? (c) Discuss the effects this could have for a compass on the submarine that is not shielded.

Solution:

(a) $4.80 \times 10^{-4} \text{ T}$

(b) Zero

(c) If the wires are not paired, the field is about 10 times stronger than Earth's magnetic field and so could severely disrupt the use of a compass.

Exercise:**Problem:**

How strong is the magnetic field inside a solenoid with 10,000 turns per meter that carries 20.0 A?

Exercise:**Problem:**

What current is needed in the solenoid described in [\[link\]](#) to produce a magnetic field 10^4 times the Earth's magnetic field of $5.00 \times 10^{-5} \text{ T}$?

Solution:

39.8 A

Exercise:

Problem:

How far from the starter cable of a car, carrying 150 A, must you be to experience a field less than the Earth's ($5.00 \times 10^{-5} \text{ T}$)? Assume a long straight wire carries the current. (In practice, the body of your car shields the dashboard compass.)

Exercise:**Problem:**

Measurements affect the system being measured, such as the current loop in [\[link\]](#). (a) Estimate the field the loop creates by calculating the field at the center of a circular loop 20.0 cm in diameter carrying 5.00 A. (b) What is the smallest field strength this loop can be used to measure, if its field must alter the measured field by less than 0.0100%?

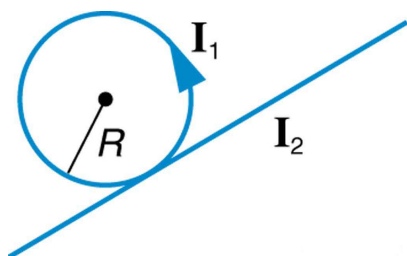
Solution:

(a) $3.14 \times 10^{-5} \text{ T}$

(b) 0.314 T

Exercise:**Problem:**

[\[link\]](#) shows a long straight wire just touching a loop carrying a current I_1 . Both lie in the same plane. (a) What direction must the current I_2 in the straight wire have to create a field at the center of the loop in the direction opposite to that created by the loop? (b) What is the ratio of I_1/I_2 that gives zero field strength at the center of the loop? (c) What is the direction of the field directly above the loop under this circumstance?



Exercise:

Problem:

Find the magnitude and direction of the magnetic field at the point equidistant from the wires in [\[link\]\(a\)](#), using the rules of vector addition to sum the contributions from each wire.

Solution:

$$7.55 \times 10^{-5} \text{ T}, 23.4^\circ$$

Exercise:

Problem:

Find the magnitude and direction of the magnetic field at the point equidistant from the wires in [\[link\]\(b\)](#), using the rules of vector addition to sum the contributions from each wire.

Exercise:

Problem:

What current is needed in the top wire in [\[link\]\(a\)](#) to produce a field of zero at the point equidistant from the wires, if the currents in the bottom two wires are both 10.0 A into the page?

Solution:

$$10.0 \text{ A}$$

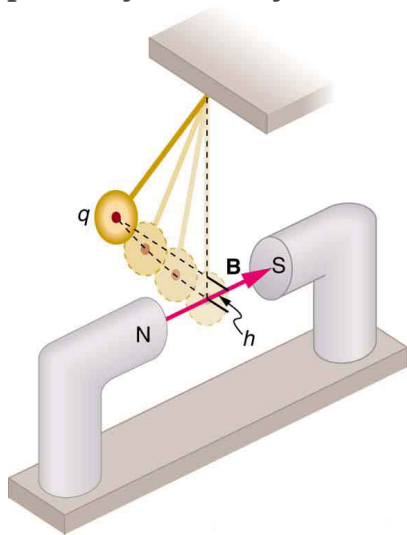
Exercise:

Problem:

Calculate the size of the magnetic field 20 m below a high voltage power line. The line carries 450 MW at a voltage of 300,000 V.

Exercise:**Problem: Integrated Concepts**

(a) A pendulum is set up so that its bob (a thin copper disk) swings between the poles of a permanent magnet as shown in [\[link\]](#). What is the magnitude and direction of the magnetic force on the bob at the lowest point in its path, if it has a positive $0.250\ \mu\text{C}$ charge and is released from a height of 30.0 cm above its lowest point? The magnetic field strength is 1.50 T. (b) What is the acceleration of the bob at the bottom of its swing if its mass is 30.0 grams and it is hung from a flexible string? Be certain to include a free-body diagram as part of your analysis.

**Solution:**

(a) $9.09 \times 10^{-7}\ \text{N}$ upward

(b) $3.03 \times 10^{-5}\ \text{m/s}^2$

Exercise:

Problem: Integrated Concepts

(a) What voltage will accelerate electrons to a speed of $6.00 \times 10^{-7} \text{ m/s}$? (b) Find the radius of curvature of the path of a *proton* accelerated through this potential in a 0.500-T field and compare this with the radius of curvature of an electron accelerated through the same potential.

Exercise:**Problem: Integrated Concepts**

Find the radius of curvature of the path of a 25.0-MeV proton moving perpendicularly to the 1.20-T field of a cyclotron.

Solution:

60.2 cm

Exercise:**Problem: Integrated Concepts**

To construct a nonmechanical water meter, a 0.500-T magnetic field is placed across the supply water pipe to a home and the Hall voltage is recorded. (a) Find the flow rate in liters per second through a 3.00-cm-diameter pipe if the Hall voltage is 60.0 mV. (b) What would the Hall voltage be for the same flow rate through a 10.0-cm-diameter pipe with the same field applied?

Exercise:**Problem: Integrated Concepts**

(a) Using the values given for an MHD drive in [\[link\]](#), and assuming the force is uniformly applied to the fluid, calculate the pressure created in N/m^2 . (b) Is this a significant fraction of an atmosphere?

Solution:

(a) $1.02 \times 10^3 \text{ N/m}^2$

(b) Not a significant fraction of an atmosphere

Exercise:**Problem: Integrated Concepts**

(a) Calculate the maximum torque on a 50-turn, 1.50 cm radius circular current loop carrying 50 μA in a 0.500-T field. (b) If this coil is to be used in a galvanometer that reads 50 μA full scale, what force constant spring must be used, if it is attached 1.00 cm from the axis of rotation and is stretched by the 60° arc moved?

Exercise:**Problem: Integrated Concepts**

A current balance used to define the ampere is designed so that the current through it is constant, as is the distance between wires. Even so, if the wires change length with temperature, the force between them will change. What percent change in force per degree will occur if the wires are copper?

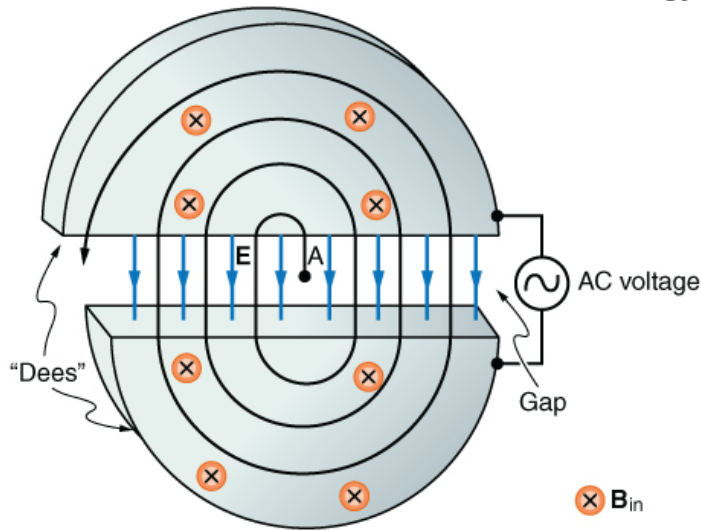
Solution:

$$17.0 \times 10^{-4} \% / ^\circ\text{C}$$

Exercise:**Problem: Integrated Concepts**

(a) Show that the period of the circular orbit of a charged particle moving perpendicularly to a uniform magnetic field is $T = 2\pi m / (qB)$. (b) What is the frequency f ? (c) What is the angular

velocity ω ? Note that these results are independent of the velocity and radius of the orbit and, hence, of the energy of the particle. ([link](#).)



Cyclotrons accelerate charged particles orbiting in a magnetic field by placing an AC voltage on the metal Dees, between which the particles move, so that energy is added twice each orbit. The frequency is constant, since it is independent of the particle energy—the radius of the orbit simply increases with energy until the particles approach the edge and are extracted for various experiments and applications.

Exercise:

Problem: Integrated Concepts

A cyclotron accelerates charged particles as shown in [link](#). Using the results of the previous problem, calculate the frequency of the accelerating voltage needed for a proton in a 1.20-T field.

Solution:

18.3 MHz

Exercise:**Problem: Integrated Concepts**

(a) A 0.140-kg baseball, pitched at 40.0 m/s horizontally and perpendicular to the Earth's horizontal 5.00×10^{-5} T field, has a 100-nC charge on it. What distance is it deflected from its path by the magnetic force, after traveling 30.0 m horizontally? (b) Would you suggest this as a secret technique for a pitcher to throw curve balls?

Exercise:**Problem: Integrated Concepts**

(a) What is the direction of the force on a wire carrying a current due east in a location where the Earth's field is due north? Both are parallel to the ground. (b) Calculate the force per meter if the wire carries 20.0 A and the field strength is 3.00×10^{-5} T. (c) What diameter copper wire would have its weight supported by this force? (d) Calculate the resistance per meter and the voltage per meter needed.

Solution:

(a) Straight up

(b) 6.00×10^{-4} N/m

(c) 94.1 μ m

(d) 2.47 Ω /m, 49.4 V/m

Exercise:**Problem: Integrated Concepts**

One long straight wire is to be held directly above another by repulsion between their currents. The lower wire carries 100 A and the wire 7.50 cm above it is 10-gauge (2.588 mm diameter) copper wire. (a) What current must flow in the upper wire, neglecting the Earth's field? (b) What is the smallest current if the Earth's 3.00×10^{-5} T field is parallel to the ground and is not neglected? (c) Is the supported wire in a stable or unstable equilibrium if displaced vertically? If displaced horizontally?

Exercise:

Problem: Unreasonable Results

(a) Find the charge on a baseball, thrown at 35.0 m/s perpendicular to the Earth's 5.00×10^{-5} T field, that experiences a 1.00-N magnetic force. (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 571 C

(b) Impossible to have such a large separated charge on such a small object.

(c) The 1.00-N force is much too great to be realistic in the Earth's field.

Exercise:

Problem: Unreasonable Results

A charged particle having mass 6.64×10^{-27} kg (that of a helium atom) moving at 8.70×10^5 m/s perpendicular to a 1.50-T magnetic field travels in a circular path of radius 16.0 mm. (a) What is the charge of the particle? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

An inventor wants to generate 120-V power by moving a 1.00-m-long wire perpendicular to the Earth's 5.00×10^{-5} T field. (a) Find the speed with which the wire must move. (b) What is unreasonable about this result? (c) Which assumption is responsible?

Solution:

(a) 2.40×10^6 m/s

(b) The speed is too high to be practical $\leq 1\%$ speed of light

(c) The assumption that you could reasonably generate such a voltage with a single wire in the Earth's field is unreasonable

Exercise:**Problem: Unreasonable Results**

Frustrated by the small Hall voltage obtained in blood flow measurements, a medical physicist decides to increase the applied magnetic field strength to get a 0.500-V output for blood moving at 30.0 cm/s in a 1.50-cm-diameter vessel. (a) What magnetic field strength is needed? (b) What is unreasonable about this result? (c) Which premise is responsible?

Exercise:**Problem: Unreasonable Results**

A surveyor 100 m from a long straight 200-kV DC power line suspects that its magnetic field may equal that of the Earth and affect compass readings. (a) Calculate the current in the wire needed to create a 5.00×10^{-5} T field at this distance. (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 25.0 kA

(b) This current is unreasonably high. It implies a total power delivery in the line of 50.0×10^9 W, which is much too high for standard transmission lines.

(c) 100 meters is a long distance to obtain the required field strength. Also coaxial cables are used for transmission lines so that there is virtually no field for DC power lines, because of cancellation from opposing currents. The surveyor's concerns are not a problem for his magnetic field measurements.

Exercise:**Problem: Construct Your Own Problem**

Consider a mass separator that applies a magnetic field perpendicular to the velocity of ions and separates the ions based on the radius of curvature of their paths in the field. Construct a problem in which you calculate the magnetic field strength needed to separate two ions that differ in mass, but not charge, and have the same initial velocity. Among the things to consider are the types of ions, the velocities they can be given before entering the magnetic field, and a reasonable value for the radius of curvature of the paths they follow. In addition, calculate the separation distance between the ions at the point where they are detected.

Exercise:**Problem: Construct Your Own Problem**

Consider using the torque on a current-carrying coil in a magnetic field to detect relatively small magnetic fields (less than the field of the Earth, for example). Construct a problem in which you calculate the maximum torque on a current-carrying loop in a magnetic field. Among the things to be considered are the size of the coil, the number

of loops it has, the current you pass through the coil, and the size of the field you wish to detect. Discuss whether the torque produced is large enough to be effectively measured. Your instructor may also wish for you to consider the effects, if any, of the field produced by the coil on the surroundings that could affect detection of the small field.

Glossary

magnetic resonance imaging (MRI)

a medical imaging technique that uses magnetic fields create detailed images of internal tissues and organs

nuclear magnetic resonance (NMR)

a phenomenon in which an externally applied magnetic field interacts with the nuclei of certain atoms

magnetocardiogram (MCG)

a recording of the heart's magnetic field as it beats

magnetoencephalogram (MEG)

a measurement of the brain's magnetic field

Introduction to Electromagnetic Waves

class="introduction"

Human eyes
detect these
orange “sea
goldie” fish
swimming
over a coral
reef in the
blue waters
of the Gulf
of Eilat (Red
Sea) using
visible light.

(credit:
Daviddarom
, Wikimedia
Commons)



The beauty of a coral reef, the warm radiance of sunshine, the sting of sunburn, the X-ray revealing a broken bone, even microwave popcorn—all are brought to us by **electromagnetic waves**. The list of the various types of electromagnetic waves, ranging from radio transmission waves to nuclear gamma-ray (γ -ray) emissions, is interesting in itself.

Even more intriguing is that all of these widely varied phenomena are different manifestations of the same thing—electromagnetic waves. (See [\[link\]](#).) What are electromagnetic waves? How are they created, and how do they travel? How can we understand and organize their widely varying properties? What is their relationship to electric and magnetic effects? These and other questions will be explored.

Note:**Misconception Alert: Sound Waves vs. Radio Waves**

Many people confuse sound waves with **radio waves**, one type of electromagnetic (EM) wave. However, sound and radio waves are

completely different phenomena. Sound creates pressure variations (waves) in matter, such as air or water, or your eardrum. Conversely, radio waves are *electromagnetic waves*, like visible light, infrared, ultraviolet, X-rays, and gamma rays. EM waves don't need a medium in which to propagate; they can travel through a vacuum, such as outer space. A radio works because sound waves played by the D.J. at the radio station are converted into electromagnetic waves, then encoded and transmitted in the radio-frequency range. The radio in your car receives the radio waves, decodes the information, and uses a speaker to change it back into a sound wave, bringing sweet music to your ears.

Discovering a New Phenomenon

It is worth noting at the outset that the general phenomenon of electromagnetic waves was predicted by theory before it was realized that light is a form of electromagnetic wave. The prediction was made by James Clerk Maxwell in the mid-19th century when he formulated a single theory combining all the electric and magnetic effects known by scientists at that time. "Electromagnetic waves" was the name he gave to the phenomena his theory predicted.

Such a theoretical prediction followed by experimental verification is an indication of the power of science in general, and physics in particular. The underlying connections and unity of physics allow certain great minds to solve puzzles without having all the pieces. The prediction of electromagnetic waves is one of the most spectacular examples of this power. Certain others, such as the prediction of antimatter, will be discussed in later modules.



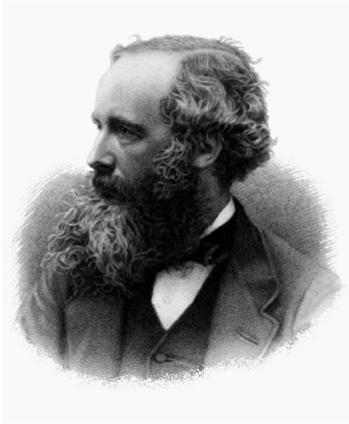
The
electromagnetic
waves sent
and received by
this 50-foot
radar dish
antenna at
Kennedy Space
Center in
Florida are not
visible, but
help track
expendable
launch vehicles
with high-
definition
imagery. The
first use of this
C-band radar
dish was for
the launch of
the Atlas V
rocket sending
the New
Horizons probe

toward Pluto.
(credit: NASA)

Maxwell's Equations: Electromagnetic Waves Predicted and Observed

- Restate Maxwell's equations.

The Scotsman James Clerk Maxwell (1831–1879) is regarded as the greatest theoretical physicist of the 19th century. (See [\[link\]](#).) Although he died young, Maxwell not only formulated a complete electromagnetic theory, represented by **Maxwell's equations**, he also developed the kinetic theory of gases and made significant contributions to the understanding of color vision and the nature of Saturn's rings.



James Clerk Maxwell, a 19th-century physicist, developed a theory that explained the relationship between electricity and magnetism and correctly predicted that visible light is caused by electromagnetic

waves. (credit:
G. J. Stodart)

Maxwell brought together all the work that had been done by brilliant physicists such as Oersted, Coulomb, Gauss, and Faraday, and added his own insights to develop the overarching theory of electromagnetism. Maxwell's equations are paraphrased here in words because their mathematical statement is beyond the level of this text. However, the equations illustrate how apparently simple mathematical statements can elegantly unite and express a multitude of concepts—why mathematics is the language of science.

Note:

Maxwell's Equations

1. **Electric field lines** originate on positive charges and terminate on negative charges. The electric field is defined as the force per unit charge on a test charge, and the strength of the force is related to the electric constant ϵ_0 , also known as the permittivity of free space. From Maxwell's first equation we obtain a special form of Coulomb's law known as Gauss's law for electricity.
2. **Magnetic field lines** are continuous, having no beginning or end. No magnetic monopoles are known to exist. The strength of the magnetic force is related to the magnetic constant μ_0 , also known as the permeability of free space. This second of Maxwell's equations is known as Gauss's law for magnetism.
3. A changing magnetic field induces an electromotive force (emf) and, hence, an electric field. The direction of the emf opposes the change. This third of Maxwell's equations is Faraday's law of induction, and includes Lenz's law.
4. Magnetic fields are generated by moving charges or by changing electric fields. This fourth of Maxwell's equations encompasses Ampere's law and adds another source of magnetism—changing electric fields.

Maxwell's equations encompass the major laws of electricity and magnetism. What is not so apparent is the symmetry that Maxwell introduced in his mathematical framework. Especially important is his addition of the hypothesis that changing electric fields create magnetic fields. This is exactly analogous (and symmetric) to Faraday's law of induction and had been suspected for some time, but fits beautifully into Maxwell's equations.

Symmetry is apparent in nature in a wide range of situations. In contemporary research, symmetry plays a major part in the search for subatomic particles using massive multinational particle accelerators such as the new Large Hadron Collider at CERN.

Note:

Making Connections: Unification of Forces

Maxwell's complete and symmetric theory showed that electric and magnetic forces are not separate, but different manifestations of the same thing—the electromagnetic force. This classical unification of forces is one motivation for current attempts to unify the four basic forces in nature—the gravitational, electrical, strong, and weak nuclear forces.

Since changing electric fields create relatively weak magnetic fields, they could not be easily detected at the time of Maxwell's hypothesis. Maxwell realized, however, that oscillating charges, like those in AC circuits, produce changing electric fields. He predicted that these changing fields would propagate from the source like waves generated on a lake by a jumping fish.

The waves predicted by Maxwell would consist of oscillating electric and magnetic fields—defined to be an electromagnetic wave (EM wave). Electromagnetic waves would be capable of exerting forces on charges great distances from their source, and they might thus be detectable. Maxwell calculated that electromagnetic waves would propagate at a speed given by the equation

Equation:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}.$$

When the values for μ_0 and ϵ_0 are entered into the equation for c , we find that

Equation:

$$c = \frac{1}{\sqrt{(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{N} \cdot \text{m}^2})(4\pi \times 10^{-7} \frac{\text{T} \cdot \text{m}}{\text{A}})}} = 3.00 \times 10^8 \text{ m/s},$$

which is the speed of light. In fact, Maxwell concluded that light is an electromagnetic wave having such wavelengths that it can be detected by the eye.

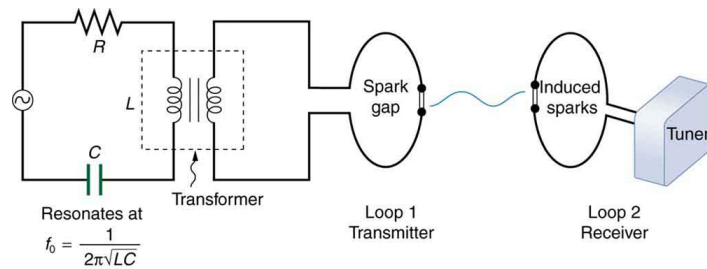
Other wavelengths should exist—it remained to be seen if they did. If so, Maxwell's theory and remarkable predictions would be verified, the greatest triumph of physics since Newton. Experimental verification came within a few years, but not before Maxwell's death.

Hertz's Observations

The German physicist Heinrich Hertz (1857–1894) was the first to generate and detect certain types of electromagnetic waves in the laboratory. Starting in 1887, he performed a series of experiments that not only confirmed the existence of electromagnetic waves, but also verified that they travel at the speed of light.

Hertz used an AC RLC (resistor-inductor-capacitor) circuit that resonates at a known frequency $f_0 = \frac{1}{2\pi\sqrt{LC}}$ and connected it to a loop of wire as shown in [\[link\]](#). High voltages induced across the gap in the loop produced sparks that were visible evidence of the current in the circuit and that helped generate electromagnetic waves.

Across the laboratory, Hertz had another loop attached to another RLC circuit, which could be tuned (as the dial on a radio) to the same resonant frequency as the first and could, thus, be made to receive electromagnetic waves. This loop also had a gap across which sparks were generated, giving solid evidence that electromagnetic waves had been received.



The apparatus used by Hertz in 1887 to generate and detect electromagnetic waves. An RLC circuit connected to the first loop caused sparks across a gap in the wire loop and generated electromagnetic waves. Sparks across a gap in the second loop located across the laboratory gave evidence that the waves had been received.

Hertz also studied the reflection, refraction, and interference patterns of the electromagnetic waves he generated, verifying their wave character. He was able to determine wavelength from the interference patterns, and knowing their frequency, he could calculate the propagation speed using the equation $v = f\lambda$ (velocity—or speed—equals frequency times wavelength). Hertz was thus able to prove that electromagnetic waves travel at the speed of light. The SI unit for frequency, the hertz (1 Hz = 1 cycle/sec), is named in his honor.

Section Summary

- Electromagnetic waves consist of oscillating electric and magnetic fields and propagate at the speed of light c . They were predicted by Maxwell, who also showed that

Equation:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}},$$

where μ_0 is the permeability of free space and ϵ_0 is the permittivity of free space.

- Maxwell's prediction of electromagnetic waves resulted from his formulation of a complete and symmetric theory of electricity and magnetism, known as Maxwell's equations.
- These four equations are paraphrased in this text, rather than presented numerically, and encompass the major laws of electricity and magnetism. First is Gauss's law for electricity, second is Gauss's law for magnetism, third is Faraday's law of induction, including Lenz's law, and fourth is Ampere's law in a symmetric formulation that adds another source of magnetism—changing electric fields.

Problems & Exercises

Exercise:

Problem:

Verify that the correct value for the speed of light c is obtained when numerical values for the permeability and permittivity of free space (μ_0 and ϵ_0) are entered into the equation $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$.

Exercise:

Problem:

Show that, when SI units for μ_0 and ϵ_0 are entered, the units given by the right-hand side of the equation in the problem above are m/s.

Glossary

electromagnetic waves

radiation in the form of waves of electric and magnetic energy

Maxwell's equations

a set of four equations that comprise a complete, overarching theory of electromagnetism

RLC circuit

an electric circuit that includes a resistor, capacitor and inductor

hertz

an SI unit denoting the frequency of an electromagnetic wave, in cycles per second

speed of light

in a vacuum, such as space, the speed of light is a constant 3×10^8 m/s

electromotive force (emf)

energy produced per unit charge, drawn from a source that produces an electrical current

electric field lines

a pattern of imaginary lines that extend between an electric source and charged objects in the surrounding area, with arrows pointed away from positively charged objects and toward negatively charged objects. The more lines in the pattern, the stronger the electric field in that region

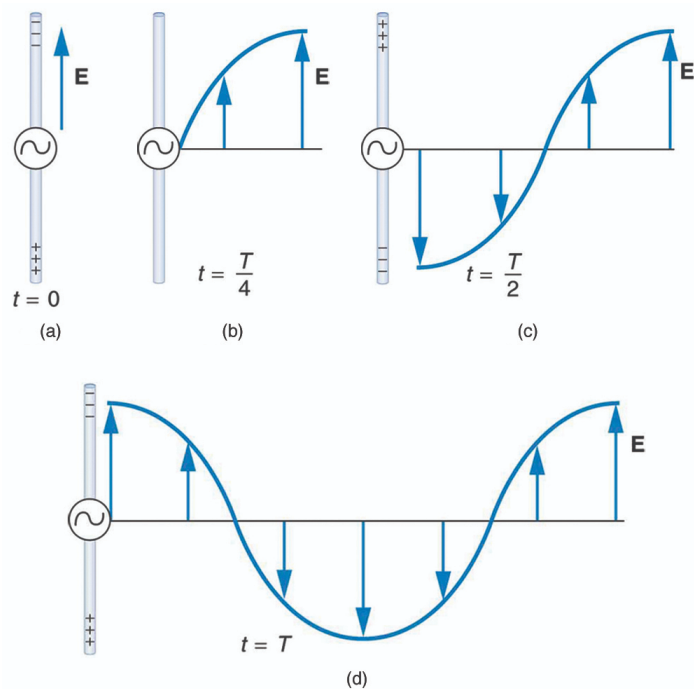
magnetic field lines

a pattern of continuous, imaginary lines that emerge from and enter into opposite magnetic poles. The density of the lines indicates the magnitude of the magnetic field

Production of Electromagnetic Waves

- Describe the electric and magnetic waves as they move out from a source, such as an AC generator.
- Explain the mathematical relationship between the magnetic field strength and the electrical field strength.
- Calculate the maximum strength of the magnetic field in an electromagnetic wave, given the maximum electric field strength.

We can get a good understanding of **electromagnetic waves** (EM) by considering how they are produced. Whenever a current varies, associated electric and magnetic fields vary, moving out from the source like waves. Perhaps the easiest situation to visualize is a varying current in a long straight wire, produced by an AC generator at its center, as illustrated in [\[link\]](#).



This long straight gray wire with an AC generator at its center becomes a broadcast antenna for electromagnetic waves. Shown here are the charge distributions at four different times. The electric field (**E**) propagates away

from the antenna at the speed of light,
forming part of an electromagnetic
wave.

The **electric field** (**E**) shown surrounding the wire is produced by the charge distribution on the wire. Both the **E** and the charge distribution vary as the current changes. The changing field propagates outward at the speed of light.

There is an associated **magnetic field** (**B**) which propagates outward as well (see [\[link\]](#)). The electric and magnetic fields are closely related and propagate as an electromagnetic wave. This is what happens in broadcast antennae such as those in radio and TV stations.

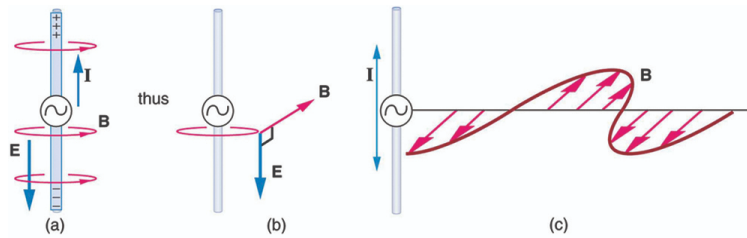
Closer examination of the one complete cycle shown in [\[link\]](#) reveals the periodic nature of the generator-driven charges oscillating up and down in the antenna and the electric field produced. At time $t = 0$, there is the maximum separation of charge, with negative charges at the top and positive charges at the bottom, producing the maximum magnitude of the electric field (or *E*-field) in the upward direction. One-fourth of a cycle later, there is no charge separation and the field next to the antenna is zero, while the maximum *E*-field has moved away at speed c .

As the process continues, the charge separation reverses and the field reaches its maximum downward value, returns to zero, and rises to its maximum upward value at the end of one complete cycle. The outgoing wave has an **amplitude** proportional to the maximum separation of charge. Its **wavelength**(λ) is proportional to the period of the oscillation and, hence, is smaller for short periods or high frequencies. (As usual, wavelength and **frequency**(f) are inversely proportional.)

Electric and Magnetic Waves: Moving Together

Following Ampere's law, current in the antenna produces a magnetic field, as shown in [\[link\]](#). The relationship between **E** and **B** is shown at one

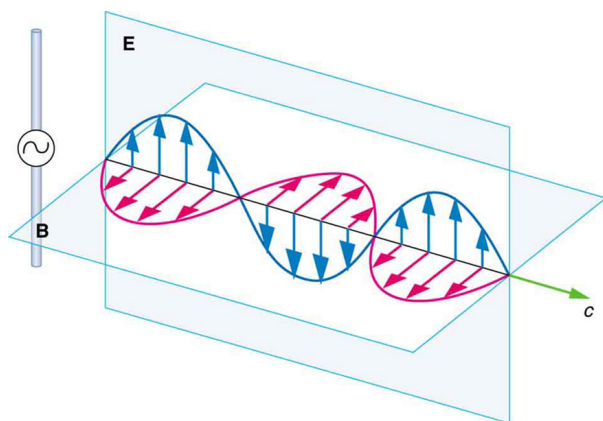
instant in [\[link\]](#) (a). As the current varies, the magnetic field varies in magnitude and direction.



(a) The current in the antenna produces the circular magnetic field lines. The current (I) produces the separation of charge along the wire, which in turn creates the electric field as shown. (b) The electric and magnetic fields (\mathbf{E} and \mathbf{B}) near the wire are perpendicular; they are shown here for one point in space. (c) The magnetic field varies with current and propagates away from the antenna at the speed of light.

The magnetic field lines also propagate away from the antenna at the speed of light, forming the other part of the electromagnetic wave, as seen in [\[link\]](#) (b). The magnetic part of the wave has the same period and wavelength as the electric part, since they are both produced by the same movement and separation of charges in the antenna.

The electric and magnetic waves are shown together at one instant in time in [\[link\]](#). The electric and magnetic fields produced by a long straight wire antenna are exactly in phase. Note that they are perpendicular to one another and to the direction of propagation, making this a **transverse wave**.



A part of the electromagnetic wave sent out from the antenna at one instant in time. The electric and magnetic fields (**E** and **B**) are in phase, and they are perpendicular to one another and the direction of propagation. For clarity, the waves are shown only along one direction, but they propagate out in other directions too.

Electromagnetic waves generally propagate out from a source in all directions, sometimes forming a complex radiation pattern. A linear antenna like this one will not radiate parallel to its length, for example. The wave is shown in one direction from the antenna in [\[link\]](#) to illustrate its basic characteristics.

Instead of the AC generator, the antenna can also be driven by an AC circuit. In fact, charges radiate whenever they are accelerated. But while a current in a circuit needs a complete path, an antenna has a varying charge distribution forming a **standing wave**, driven by the AC. The dimensions of the antenna are critical for determining the frequency of the radiated electromagnetic waves. This is a **resonant** phenomenon and when we tune

radios or TV, we vary electrical properties to achieve appropriate resonant conditions in the antenna.

Receiving Electromagnetic Waves

Electromagnetic waves carry energy away from their source, similar to a sound wave carrying energy away from a standing wave on a guitar string. An antenna for receiving EM signals works in reverse. And like antennas that produce EM waves, receiver antennas are specially designed to resonate at particular frequencies.

An incoming electromagnetic wave accelerates electrons in the antenna, setting up a standing wave. If the radio or TV is switched on, electrical components pick up and amplify the signal formed by the accelerating electrons. The signal is then converted to audio and/or video format. Sometimes big receiver dishes are used to focus the signal onto an antenna.

In fact, charges radiate whenever they are accelerated. When designing circuits, we often assume that energy does not quickly escape AC circuits, and mostly this is true. A broadcast antenna is specially designed to enhance the rate of electromagnetic radiation, and shielding is necessary to keep the radiation close to zero. Some familiar phenomena are based on the production of electromagnetic waves by varying currents. Your microwave oven, for example, sends electromagnetic waves, called microwaves, from a concealed antenna that has an oscillating current imposed on it.

Relating E -Field and B -Field Strengths

There is a relationship between the E - and B -field strengths in an electromagnetic wave. This can be understood by again considering the antenna just described. The stronger the E -field created by a separation of charge, the greater the current and, hence, the greater the B -field created.

Since current is directly proportional to voltage (Ohm's law) and voltage is directly proportional to E -field strength, the two should be directly proportional. It can be shown that the magnitudes of the fields do have a constant ratio, equal to the speed of light. That is,

Equation:

$$\frac{E}{B} = c$$

is the ratio of E -field strength to B -field strength in any electromagnetic wave. This is true at all times and at all locations in space. A simple and elegant result.

Example:**Calculating B -Field Strength in an Electromagnetic Wave**

What is the maximum strength of the B -field in an electromagnetic wave that has a maximum E -field strength of 1000 V/m?

Strategy

To find the B -field strength, we rearrange the above equation to solve for B , yielding

Equation:

$$B = \frac{E}{c}.$$

Solution

We are given E , and c is the speed of light. Entering these into the expression for B yields

Equation:

$$B = \frac{1000 \text{ V/m}}{3.00 \times 10^8 \text{ m/s}} = 3.33 \times 10^{-6} \text{ T},$$

Where T stands for Tesla, a measure of magnetic field strength.

Discussion

The B -field strength is less than a tenth of the Earth's admittedly weak magnetic field. This means that a relatively strong electric field of 1000 V/m is accompanied by a relatively weak magnetic field. Note that as this

wave spreads out, say with distance from an antenna, its field strengths become progressively weaker.

The result of this example is consistent with the statement made in the module [Maxwell's Equations: Electromagnetic Waves Predicted and Observed](#) that changing electric fields create relatively weak magnetic fields. They can be detected in electromagnetic waves, however, by taking advantage of the phenomenon of resonance, as Hertz did. A system with the same natural frequency as the electromagnetic wave can be made to oscillate. All radio and TV receivers use this principle to pick up and then amplify weak electromagnetic waves, while rejecting all others not at their resonant frequency.

Note:

Take-Home Experiment: Antennas

For your TV or radio at home, identify the antenna, and sketch its shape. If you don't have cable, you might have an outdoor or indoor TV antenna. Estimate its size. If the TV signal is between 60 and 216 MHz for basic channels, then what is the wavelength of those EM waves?

Try tuning the radio and note the small range of frequencies at which a reasonable signal for that station is received. (This is easier with digital readout.) If you have a car with a radio and extendable antenna, note the quality of reception as the length of the antenna is changed.

Note:

PhET Explorations: Radio Waves and Electromagnetic Fields

Broadcast radio waves from KPhET. Wiggle the transmitter electron manually or have it oscillate automatically. Display the field as a curve or vectors. The strip chart shows the electron positions at the transmitter and at the receiver.

<https://archive.cnx.org/specials/c8dd764c-ae74-11e5-af4c-3375261fa183/radio-waves/#sim-radio-waves>

Section Summary

- Electromagnetic waves are created by oscillating charges (which radiate whenever accelerated) and have the same frequency as the oscillation.
- Since the electric and magnetic fields in most electromagnetic waves are perpendicular to the direction in which the wave moves, it is ordinarily a transverse wave.
- The strengths of the electric and magnetic parts of the wave are related by
Equation:

$$\frac{E}{B} = c,$$

which implies that the magnetic field B is very weak relative to the electric field E .

Conceptual Questions

Exercise:

Problem:

The direction of the electric field shown in each part of [\[link\]](#) is that produced by the charge distribution in the wire. Justify the direction shown in each part, using the Coulomb force law and the definition of $\mathbf{E} = \mathbf{F}/q$, where q is a positive test charge.

Exercise:

Problem:

Is the direction of the magnetic field shown in [\[link\]](#) (a) consistent with the right-hand rule for current (RHR-2) in the direction shown in the figure?

Exercise:

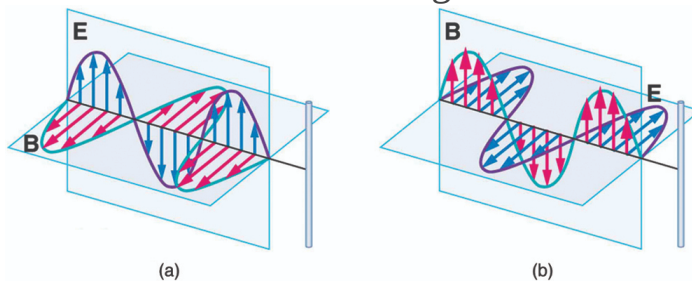
Problem:

Why is the direction of the current shown in each part of [\[link\]](#) opposite to the electric field produced by the wire's charge separation?

Exercise:

Problem:

In which situation shown in [\[link\]](#) will the electromagnetic wave be more successful in inducing a current in the wire? Explain.

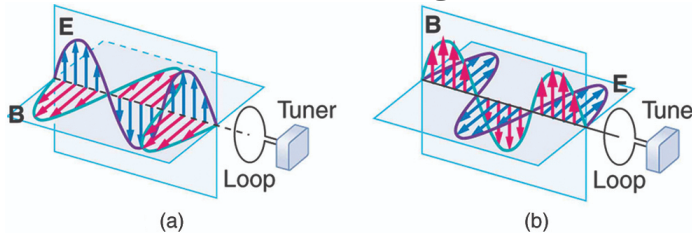


Electromagnetic waves approaching long straight wires.

Exercise:

Problem:

In which situation shown in [\[link\]](#) will the electromagnetic wave be more successful in inducing a current in the loop? Explain.



Electromagnetic waves approaching a wire loop.

Exercise:

Problem:

Should the straight wire antenna of a radio be vertical or horizontal to best receive radio waves broadcast by a vertical transmitter antenna? How should a loop antenna be aligned to best receive the signals? (Note that the direction of the loop that produces the best reception can be used to determine the location of the source. It is used for that purpose in tracking tagged animals in nature studies, for example.)

Exercise:

Problem:

Under what conditions might wires in a DC circuit emit electromagnetic waves?

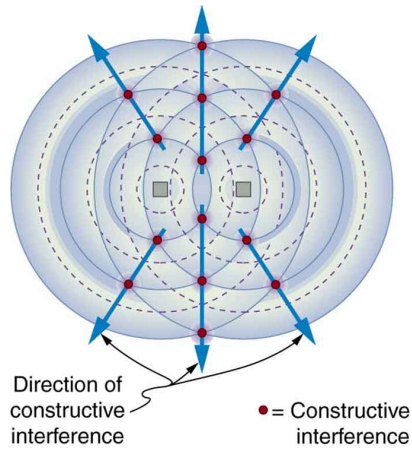
Exercise:

Problem: Give an example of interference of electromagnetic waves.

Exercise:

Problem:

[\[link\]](#) shows the interference pattern of two radio antennas broadcasting the same signal. Explain how this is analogous to the interference pattern for sound produced by two speakers. Could this be used to make a directional antenna system that broadcasts preferentially in certain directions? Explain.



An overhead view
of two radio
broadcast antennas
sending the same
signal, and the
interference pattern
they produce.

Exercise:

Problem: Can an antenna be any length? Explain your answer.

Problems & Exercises

Exercise:

Problem:

What is the maximum electric field strength in an electromagnetic wave that has a maximum magnetic field strength of $5.00 \times 10^{-4} \text{ T}$ (about 10 times the Earth's)?

Solution:

150 kV/m

Exercise:

Problem:

The maximum magnetic field strength of an electromagnetic field is 5×10^{-6} T. Calculate the maximum electric field strength if the wave is traveling in a medium in which the speed of the wave is $0.75c$.

Exercise:

Problem:

Verify the units obtained for magnetic field strength B in [\[link\]](#) (using the equation $B = \frac{E}{c}$) are in fact teslas (T).

Glossary

electric field

a vector quantity (**E**); the lines of electric force per unit charge, moving radially outward from a positive charge and in toward a negative charge

electric field strength

the magnitude of the electric field, denoted E -field

magnetic field

a vector quantity (**B**); can be used to determine the magnetic force on a moving charged particle

magnetic field strength

the magnitude of the magnetic field, denoted B -field

transverse wave

a wave, such as an electromagnetic wave, which oscillates perpendicular to the axis along the line of travel

standing wave

a wave that oscillates in place, with nodes where no motion happens

wavelength

the distance from one peak to the next in a wave

amplitude

the height, or magnitude, of an electromagnetic wave

frequency

the number of complete wave cycles (up-down-up) passing a given point within one second (cycles/second)

resonant

a system that displays enhanced oscillation when subjected to a periodic disturbance of the same frequency as its natural frequency

oscillate

to fluctuate back and forth in a steady beat

The Electromagnetic Spectrum

- List three “rules of thumb” that apply to the different frequencies along the electromagnetic spectrum.
- Explain why the higher the frequency, the shorter the wavelength of an electromagnetic wave.
- Draw a simplified electromagnetic spectrum, indicating the relative positions, frequencies, and spacing of the different types of radiation bands.
- List and explain the different methods by which electromagnetic waves are produced across the spectrum.

In this module we examine how electromagnetic waves are classified into categories such as radio, infrared, ultraviolet, and so on, so that we can understand some of their similarities as well as some of their differences. We will also find that there are many connections with previously discussed topics, such as wavelength and resonance. A brief overview of the production and utilization of electromagnetic waves is found in [\[link\]](#).

Type of EM wave	Production	Applications	Life sciences aspect	Issues
Radio & TV	Accelerating charges	Communications Remote controls	MRI	Requires controls for band use
Microwaves	Accelerating charges & thermal agitation	Communications Ovens Radar	Deep heating	Cell phone use
Infrared	Thermal agitations & electronic transitions	Thermal imaging Heating	Absorbed by atmosphere	Greenhouse effect
Visible light	Thermal agitations & electronic transitions	All pervasive	Photosynthesis Human vision	

Type of EM wave	Production	Applications	Life sciences aspect	Issues
Ultraviolet	Thermal agitations & electronic transitions	Sterilization Cancer control	Vitamin D production	Ozone depletion Cancer causing
X-rays	Inner electronic transitions and fast collisions	Medical Security	Medical diagnosis Cancer therapy	Cancer causing
Gamma rays	Nuclear decay	Nuclear medicineSecurity	Medical diagnosis Cancer therapy	Cancer causing Radiation damage

Electromagnetic Waves

Note:

Connections: Waves

There are many types of waves, such as water waves and even earthquakes. Among the many shared attributes of waves are propagation speed, frequency, and wavelength. These are always related by the expression $v_W = f\lambda$. This module concentrates on EM waves, but other modules contain examples of all of these characteristics for sound waves and submicroscopic particles.

As noted before, an electromagnetic wave has a frequency and a wavelength associated with it and travels at the speed of light, or c . The relationship among these wave characteristics can be described by $v_W = f\lambda$, where v_W is the propagation speed of the wave, f is the frequency, and λ is the wavelength. Here $v_W = c$, so that for all electromagnetic waves,

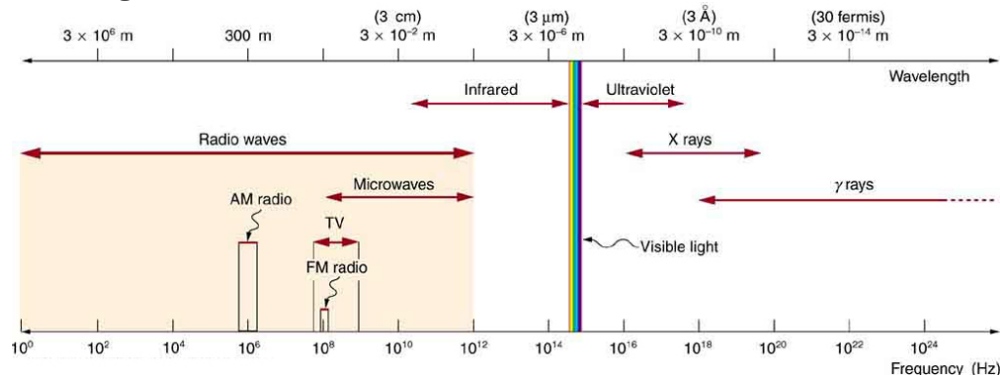
Equation:

$$c = f\lambda.$$

Thus, for all electromagnetic waves, the greater the frequency, the smaller the wavelength.

[\[link\]](#) shows how the various types of electromagnetic waves are categorized according to their wavelengths and frequencies—that is, it shows the electromagnetic spectrum. Many of the

characteristics of the various types of electromagnetic waves are related to their frequencies and wavelengths, as we shall see.



The electromagnetic spectrum, showing the major categories of electromagnetic waves. The range of frequencies and wavelengths is remarkable. The dividing line between some categories is distinct, whereas other categories overlap.

Note:

Electromagnetic Spectrum: Rules of Thumb

Three rules that apply to electromagnetic waves in general are as follows:

- High-frequency electromagnetic waves are more energetic and are more able to penetrate than low-frequency waves.
- High-frequency electromagnetic waves can carry more information per unit time than low-frequency waves.
- The shorter the wavelength of any electromagnetic wave probing a material, the smaller the detail it is possible to resolve.

Note that there are exceptions to these rules of thumb.

Transmission, Reflection, and Absorption

What happens when an electromagnetic wave impinges on a material? If the material is transparent to the particular frequency, then the wave can largely be transmitted. If the material is opaque to the frequency, then the wave can be totally reflected. The wave can also be absorbed by the material, indicating that there is some interaction between the wave and the material, such as the thermal agitation of molecules.

Of course it is possible to have partial transmission, reflection, and absorption. We normally associate these properties with visible light, but they do apply to all electromagnetic waves.

What is not obvious is that something that is transparent to light may be opaque at other frequencies. For example, ordinary glass is transparent to visible light but largely opaque to ultraviolet radiation. Human skin is opaque to visible light—we cannot see through people—but transparent to X-rays.

Radio and TV Waves

The broad category of **radio waves** is defined to contain any electromagnetic wave produced by currents in wires and circuits. Its name derives from their most common use as a carrier of audio information (i.e., radio). The name is applied to electromagnetic waves of similar frequencies regardless of source. Radio waves from outer space, for example, do not come from alien radio stations. They are created by many astronomical phenomena, and their study has revealed much about nature on the largest scales.

There are many uses for radio waves, and so the category is divided into many subcategories, including microwaves and those electromagnetic waves used for AM and FM radio, cellular telephones, and TV.

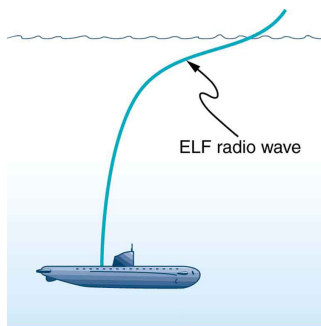
The lowest commonly encountered radio frequencies are produced by high-voltage AC power transmission lines at frequencies of 50 or 60 Hz. (See [\[link\]](#).) These extremely long wavelength electromagnetic waves (about 6000 km!) are one means of energy loss in long-distance power transmission.



This high-voltage traction power line running to Eutingen Railway Substation in Germany radiates electromagnetic waves with very long wavelengths. (credit: Zonk43, Wikimedia Commons)

There is an ongoing controversy regarding potential health hazards associated with exposure to these electromagnetic fields (*E*-fields). Some people suspect that living near such transmission lines may cause a variety of illnesses, including cancer. But demographic data are either inconclusive or simply do not support the hazard theory. Recent reports that have looked at many European and American epidemiological studies have found no increase in risk for cancer due to exposure to *E*-fields.

Extremely low frequency (ELF) radio waves of about 1 kHz are used to communicate with submerged submarines. The ability of radio waves to penetrate salt water is related to their wavelength (much like ultrasound penetrating tissue)—the longer the wavelength, the farther they penetrate. Since salt water is a good conductor, radio waves are strongly absorbed by it, and very long wavelengths are needed to reach a submarine under the surface. (See [\[link\]](#).)

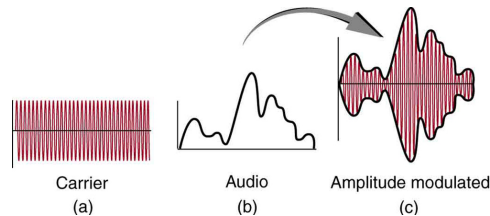


Very long wavelength radio waves are needed to reach this submarine, requiring extremely low frequency signals (ELF). Shorter wavelengths do not penetrate to any significant depth.

AM radio waves are used to carry commercial radio signals in the frequency range from 540 to 1600 kHz. The abbreviation AM stands for **amplitude modulation**, which is the method for placing information on these waves. (See [\[link\]](#).) A **carrier wave** having the basic frequency of the radio station, say 1530 kHz, is varied or modulated in amplitude by an audio signal. The resulting wave has a constant frequency, but a varying amplitude.

A radio receiver tuned to have the same resonant frequency as the carrier wave can pick up the signal, while rejecting the many other frequencies impinging on its antenna. The receiver's

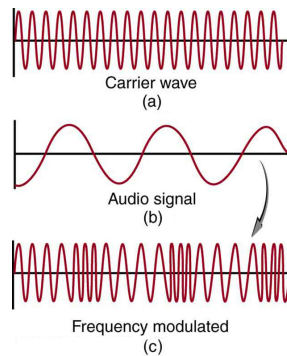
circuitry is designed to respond to variations in amplitude of the carrier wave to replicate the original audio signal. That audio signal is amplified to drive a speaker or perhaps to be recorded.



Amplitude modulation for AM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The amplitude of the carrier is modulated by the audio signal without changing its basic frequency.

FM Radio Waves

FM radio waves are also used for commercial radio transmission, but in the frequency range of 88 to 108 MHz. FM stands for **frequency modulation**, another method of carrying information. (See [\[link\]](#).) Here a carrier wave having the basic frequency of the radio station, perhaps 105.1 MHz, is modulated in frequency by the audio signal, producing a wave of constant amplitude but varying frequency.



Frequency
modulation for

FM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The frequency of the carrier is modulated by the audio signal without changing its amplitude.

Since audible frequencies range up to 20 kHz (or 0.020 MHz) at most, the frequency of the FM radio wave can vary from the carrier by as much as 0.020 MHz. Thus the carrier frequencies of two different radio stations cannot be closer than 0.020 MHz. An FM receiver is tuned to resonate at the carrier frequency and has circuitry that responds to variations in frequency, reproducing the audio information.

FM radio is inherently less subject to noise from stray radio sources than AM radio. The reason is that amplitudes of waves add. So an AM receiver would interpret noise added onto the amplitude of its carrier wave as part of the information. An FM receiver can be made to reject amplitudes other than that of the basic carrier wave and only look for variations in frequency. It is thus easier to reject noise from FM, since noise produces a variation in amplitude.

Television is also broadcast on electromagnetic waves. Since the waves must carry a great deal of visual as well as audio information, each channel requires a larger range of frequencies than simple radio transmission. TV channels utilize frequencies in the range of 54 to 88 MHz and 174 to 222 MHz. (The entire FM radio band lies between channels 88 MHz and 174 MHz.) These TV channels are called VHF (for **very high frequency**). Other channels called UHF (for **ultra high frequency**) utilize an even higher frequency range of 470 to 1000 MHz.

The TV video signal is AM, while the TV audio is FM. Note that these frequencies are those of free transmission with the user utilizing an old-fashioned roof antenna. Satellite dishes and cable transmission of TV occurs at significantly higher frequencies and is rapidly evolving with the use of the high-definition or HD format.

Example: **Calculating Wavelengths of Radio Waves**

Calculate the wavelengths of a 1530-kHz AM radio signal, a 105.1-MHz FM radio signal, and a 1.90-GHz cell phone signal.

Strategy

The relationship between wavelength and frequency is $c = f\lambda$, where $c = 3.00 \times 10^8$ m/s is the speed of light (the speed of light is only very slightly smaller in air than it is in a vacuum). We can rearrange this equation to find the wavelength for all three frequencies.

Solution

Rearranging gives

Equation:

$$\lambda = \frac{c}{f}.$$

(a) For the $f = 1530$ kHz AM radio signal, then,

Equation:

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{1530 \times 10^3 \text{ cycles/s}} \\ &= 196 \text{ m.}\end{aligned}$$

(b) For the $f = 105.1$ MHz FM radio signal,

Equation:

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{105.1 \times 10^6 \text{ cycles/s}} \\ &= 2.85 \text{ m.}\end{aligned}$$

(c) And for the $f = 1.90$ GHz cell phone,

Equation:

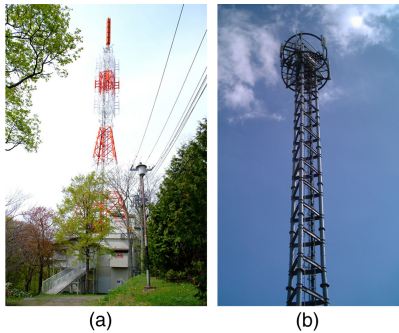
$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{1.90 \times 10^9 \text{ cycles/s}} \\ &= 0.158 \text{ m.}\end{aligned}$$

Discussion

These wavelengths are consistent with the spectrum in [\[link\]](#). The wavelengths are also related to other properties of these electromagnetic waves, as we shall see.

The wavelengths found in the preceding example are representative of AM, FM, and cell phones, and account for some of the differences in how they are broadcast and how well they travel. The most efficient length for a linear antenna, such as discussed in [Production of Electromagnetic Waves](#), is $\lambda/2$, half the wavelength of the electromagnetic wave. Thus a very large antenna is needed to efficiently broadcast typical AM radio with its carrier wavelengths on the order of hundreds of meters.

One benefit to these long AM wavelengths is that they can go over and around rather large obstacles (like buildings and hills), just as ocean waves can go around large rocks. FM and TV are best received when there is a line of sight between the broadcast antenna and receiver, and they are often sent from very tall structures. FM, TV, and mobile phone antennas themselves are much smaller than those used for AM, but they are elevated to achieve an unobstructed line of sight. (See [\[link\]](#).)



(a) A large tower is used to broadcast TV signals.

The actual antennas are small structures on top of the tower—they are placed at great heights to have a clear line of sight over a large broadcast area. (credit: Ozizo, Wikimedia Commons)

(b) The NTT Dokomo mobile phone tower at Tokorozawa City, Japan.

(credit: tokoroten, Wikimedia Commons)

Radio Wave Interference

Astronomers and astrophysicists collect signals from outer space using electromagnetic waves. A common problem for astrophysicists is the “pollution” from electromagnetic radiation pervading our surroundings from communication systems in general. Even everyday gadgets like our car keys having the facility to lock car doors remotely and being able to turn TVs on and off using remotes involve radio-wave frequencies. In order to prevent interference between all these electromagnetic signals, strict regulations are drawn up for different organizations to utilize different radio frequency bands.

One reason why we are sometimes asked to switch off our mobile phones (operating in the range of 1.9 GHz) on airplanes and in hospitals is that important communications or medical equipment often uses similar radio frequencies and their operation can be affected by frequencies used in the communication devices.

For example, radio waves used in magnetic resonance imaging (MRI) have frequencies on the order of 100 MHz, although this varies significantly depending on the strength of the magnetic field used and the nuclear type being scanned. MRI is an important medical imaging and research tool, producing highly detailed two- and three-dimensional images. Radio waves are broadcast, absorbed, and reemitted in a resonance process that is sensitive to the density of nuclei (usually protons or hydrogen nuclei).

The wavelength of 100-MHz radio waves is 3 m, yet using the sensitivity of the resonant frequency to the magnetic field strength, details smaller than a millimeter can be imaged. This is a good example of an exception to a rule of thumb (in this case, the rubric that details much smaller than the probe's wavelength cannot be detected). The intensity of the radio waves used in MRI presents little or no hazard to human health.

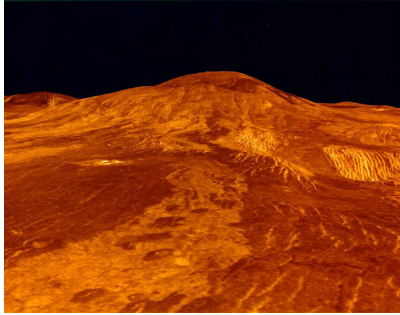
Microwaves

Microwaves are the highest-frequency electromagnetic waves that can be produced by currents in macroscopic circuits and devices. Microwave frequencies range from about 10^9 Hz to the highest practical LC resonance at nearly 10^{12} Hz. Since they have high frequencies, their wavelengths are short compared with those of other radio waves—hence the name “microwave.”

Microwaves can also be produced by atoms and molecules. They are, for example, a component of electromagnetic radiation generated by **thermal agitation**. The thermal motion of atoms and molecules in any object at a temperature above absolute zero causes them to emit and absorb radiation.

Since it is possible to carry more information per unit time on high frequencies, microwaves are quite suitable for communications. Most satellite-transmitted information is carried on microwaves, as are land-based long-distance transmissions. A clear line of sight between transmitter and receiver is needed because of the short wavelengths involved.

Radar is a common application of microwaves that was first developed in World War II. By detecting and timing microwave echoes, radar systems can determine the distance to objects as diverse as clouds and aircraft. A Doppler shift in the radar echo can be used to determine the speed of a car or the intensity of a rainstorm. Sophisticated radar systems are used to map the Earth and other planets, with a resolution limited by wavelength. (See [\[link\]](#).) The shorter the wavelength of any probe, the smaller the detail it is possible to observe.



An image of Sif Mons with lava flows on Venus, based on Magellan synthetic aperture radar data combined with radar altimetry to produce a three-dimensional map of the surface. The Venusian atmosphere is opaque to visible light, but not to the microwaves that were used to create this image.

(credit: NSSDC, NASA/JPL)

Heating with Microwaves

How does the ubiquitous microwave oven produce microwaves electronically, and why does food absorb them preferentially? Microwaves at a frequency of 2.45 GHz are produced by accelerating electrons. The microwaves are then used to induce an alternating electric field in the oven.

Water and some other constituents of food have a slightly negative charge at one end and a slightly positive charge at one end (called polar molecules). The range of microwave frequencies is specially selected so that the polar molecules, in trying to keep orienting themselves with the electric field, absorb these energies and increase their temperatures—called dielectric heating.

The energy thereby absorbed results in thermal agitation heating food and not the plate, which does not contain water. Hot spots in the food are related to constructive and destructive interference patterns. Rotating antennas and food turntables help spread out the hot spots.

Another use of microwaves for heating is within the human body. Microwaves will penetrate more than shorter wavelengths into tissue and so can accomplish “deep heating” (called

microwave diathermy). This is used for treating muscular pains, spasms, tendonitis, and rheumatoid arthritis.

Note:

Making Connections: Take-Home Experiment—Microwave Ovens

1. Look at the door of a microwave oven. Describe the structure of the door. Why is there a metal grid on the door? How does the size of the holes in the grid compare with the wavelengths of microwaves used in microwave ovens? What is this wavelength?
2. Place a glass of water (about 250 ml) in the microwave and heat it for 30 seconds. Measure the temperature gain (the ΔT). Assuming that the power output of the oven is 1000 W, calculate the efficiency of the heat-transfer process.
3. Remove the rotating turntable or moving plate and place a cup of water in several places along a line parallel with the opening. Heat for 30 seconds and measure the ΔT for each position. Do you see cases of destructive interference?

Microwaves generated by atoms and molecules far away in time and space can be received and detected by electronic circuits. Deep space acts like a blackbody with a 2.7 K temperature, radiating most of its energy in the microwave frequency range. In 1964, Penzias and Wilson detected this radiation and eventually recognized that it was the radiation of the Big Bang's cooled remnants.

Infrared Radiation

The microwave and infrared regions of the electromagnetic spectrum overlap (see [\[link\]](#)).

Infrared radiation is generally produced by thermal motion and the vibration and rotation of atoms and molecules. Electronic transitions in atoms and molecules can also produce infrared radiation.

The range of infrared frequencies extends up to the lower limit of visible light, just below red. In fact, infrared means “below red.” Frequencies at its upper limit are too high to be produced by accelerating electrons in circuits, but small systems, such as atoms and molecules, can vibrate fast enough to produce these waves.

Water molecules rotate and vibrate particularly well at infrared frequencies, emitting and absorbing them so efficiently that the emissivity for skin is $e = 0.97$ in the infrared. Night-vision scopes can detect the infrared emitted by various warm objects, including humans, and convert it to visible light.

We can examine radiant heat transfer from a house by using a camera capable of detecting infrared radiation. Reconnaissance satellites can detect buildings, vehicles, and even individual humans by their infrared emissions, whose power radiation is proportional to the fourth power of the absolute temperature. More mundanely, we use infrared lamps, some of which are called

quartz heaters, to preferentially warm us because we absorb infrared better than our surroundings.

The Sun radiates like a nearly perfect blackbody (that is, it has $e = 1$), with a 6000 K surface temperature. About half of the solar energy arriving at the Earth is in the infrared region, with most of the rest in the visible part of the spectrum, and a relatively small amount in the ultraviolet. On average, 50 percent of the incident solar energy is absorbed by the Earth.

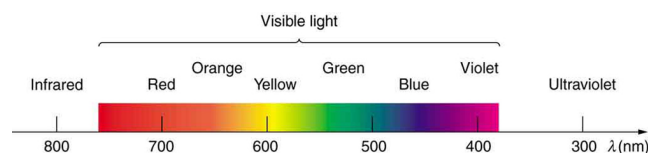
The relatively constant temperature of the Earth is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by CO_2 and H_2O in the atmosphere and then radiated back to Earth or into outer space. This radiation back to Earth is known as the greenhouse effect, and it maintains the surface temperature of the Earth about 40°C higher than it would be if there is no absorption. Some scientists think that the increased concentration of CO_2 and other greenhouse gases in the atmosphere, resulting from increases in fossil fuel burning, has increased global average temperatures.

Visible Light

Visible light is the narrow segment of the electromagnetic spectrum to which the normal human eye responds. Visible light is produced by vibrations and rotations of atoms and molecules, as well as by electronic transitions within atoms and molecules. The receivers or detectors of light largely utilize electronic transitions. We say the atoms and molecules are excited when they absorb and relax when they emit through electronic transitions.

[\[link\]](#) shows this part of the spectrum, together with the colors associated with particular pure wavelengths. We usually refer to visible light as having wavelengths of between 400 nm and 750 nm. (The retina of the eye actually responds to the lowest ultraviolet frequencies, but these do not normally reach the retina because they are absorbed by the cornea and lens of the eye.)

Red light has the lowest frequencies and longest wavelengths, while violet has the highest frequencies and shortest wavelengths. Blackbody radiation from the Sun peaks in the visible part of the spectrum but is more intense in the red than in the violet, making the Sun yellowish in appearance.



A small part of the electromagnetic spectrum that includes its visible components. The divisions between infrared, visible, and ultraviolet are not perfectly

distinct, nor are those between the seven rainbow colors.

Living things—plants and animals—have evolved to utilize and respond to parts of the electromagnetic spectrum they are embedded in. Visible light is the most predominant and we enjoy the beauty of nature through visible light. Plants are more selective. Photosynthesis makes use of parts of the visible spectrum to make sugars.

Example:

Integrated Concept Problem: Correcting Vision with Lasers

During laser vision correction, a brief burst of 193-nm ultraviolet light is projected onto the cornea of a patient. It makes a spot 0.80 mm in diameter and evaporates a layer of cornea 0.30 μm thick. Calculate the energy absorbed, assuming the corneal tissue has the same properties as water; it is initially at 34°C. Assume the evaporated tissue leaves at a temperature of 100°C.

Strategy

The energy from the laser light goes toward raising the temperature of the tissue and also toward evaporating it. Thus we have two amounts of heat to add together. Also, we need to find the mass of corneal tissue involved.

Solution

To figure out the heat required to raise the temperature of the tissue to 100°C, we can apply concepts of thermal energy. We know that

Equation:

$$Q = mc\Delta T,$$

where Q is the heat required to raise the temperature, ΔT is the desired change in temperature, m is the mass of tissue to be heated, and c is the specific heat of water equal to 4186 J/kg/K. Without knowing the mass m at this point, we have

Equation:

$$Q = m(4186 \text{ J/kg/K})(100^\circ\text{C} - 34^\circ\text{C}) = m(276,276 \text{ J/kg}) = m(276 \text{ kJ/kg}).$$

The latent heat of vaporization of water is 2256 kJ/kg, so that the energy needed to evaporate mass m is

Equation:

$$Q_v = mL_v = m(2256 \text{ kJ/kg}).$$

To find the mass m , we use the equation $\rho = m/V$, where ρ is the density of the tissue and V is its volume. For this case,

Equation:

$$\begin{aligned}
 m &= \rho V \\
 &= (1000 \text{ kg/m}^3)(\text{area} \times \text{thickness}(\text{m}^3)) \\
 &= (1000 \text{ kg/m}^3)(\pi(0.80 \times 10^{-3} \text{ m})^2/4)(0.30 \times 10^{-6} \text{ m}) \\
 &= 0.151 \times 10^{-9} \text{ kg}.
 \end{aligned}$$

Therefore, the total energy absorbed by the tissue in the eye is the sum of Q and Q_v :

Equation:

$$Q_{\text{tot}} = m(c\Delta T + L_v) = (0.151 \times 10^{-9} \text{ kg})(276 \text{ kJ/kg} + 2256 \text{ kJ/kg}) = 382 \times 10^{-9} \text{ kJ}.$$

Discussion

The lasers used for this eye surgery are excimer lasers, whose light is well absorbed by biological tissue. They evaporate rather than burn the tissue, and can be used for precision work. Most lasers used for this type of eye surgery have an average power rating of about one watt. For our example, if we assume that each laser burst from this pulsed laser lasts for 10 ns, and there are 400 bursts per second, then the average power is $Q_{\text{tot}} \times 400 = 150 \text{ mW}$.

Optics is the study of the behavior of visible light and other forms of electromagnetic waves. Optics falls into two distinct categories. When electromagnetic radiation, such as visible light, interacts with objects that are large compared with its wavelength, its motion can be represented by straight lines like rays. Ray optics is the study of such situations and includes lenses and mirrors.

When electromagnetic radiation interacts with objects about the same size as the wavelength or smaller, its wave nature becomes apparent. For example, observable detail is limited by the wavelength, and so visible light can never detect individual atoms, because they are so much smaller than its wavelength. Physical or wave optics is the study of such situations and includes all wave characteristics.

Note:

Take-Home Experiment: Colors That Match

When you light a match you see largely orange light; when you light a gas stove you see blue light. Why are the colors different? What other colors are present in these?

Ultraviolet Radiation

Ultraviolet means “above violet.” The electromagnetic frequencies of **ultraviolet radiation (UV)** extend upward from violet, the highest-frequency visible light. Ultraviolet is also produced by atomic and molecular motions and electronic transitions. The wavelengths of ultraviolet extend from 400 nm down to about 10 nm at its highest frequencies, which overlap

with the lowest X-ray frequencies. It was recognized as early as 1801 by Johann Ritter that the solar spectrum had an invisible component beyond the violet range.

Solar UV radiation is broadly subdivided into three regions: UV-A (320–400 nm), UV-B (290–320 nm), and UV-C (220–290 nm), ranked from long to shorter wavelengths (from smaller to larger energies). Most UV-B and all UV-C is absorbed by ozone (O₃) molecules in the upper atmosphere. Consequently, 99% of the solar UV radiation reaching the Earth's surface is UV-A.

Human Exposure to UV Radiation

It is largely exposure to UV-B that causes skin cancer. It is estimated that as many as 20% of adults will develop skin cancer over the course of their lifetime. Again, treatment is often successful if caught early. Despite very little UV-B reaching the Earth's surface, there are substantial increases in skin-cancer rates in countries such as Australia, indicating how important it is that UV-B and UV-C continue to be absorbed by the upper atmosphere.

All UV radiation can damage collagen fibers, resulting in an acceleration of the aging process of skin and the formation of wrinkles. Because there is so little UV-B and UV-C reaching the Earth's surface, sunburn is caused by large exposures, and skin cancer from repeated exposure. Some studies indicate a link between overexposure to the Sun when young and melanoma later in life.

The tanning response is a defense mechanism in which the body produces pigments to absorb future exposures in inert skin layers above living cells. Basically UV-B radiation excites DNA molecules, distorting the DNA helix, leading to mutations and the possible formation of cancerous cells.

Repeated exposure to UV-B may also lead to the formation of cataracts in the eyes—a cause of blindness among people living in the equatorial belt where medical treatment is limited. Cataracts, clouding in the eye's lens and a loss of vision, are age related; 60% of those between the ages of 65 and 74 will develop cataracts. However, treatment is easy and successful, as one replaces the lens of the eye with a plastic lens. Prevention is important. Eye protection from UV is more effective with plastic sunglasses than those made of glass.

A major acute effect of extreme UV exposure is the suppression of the immune system, both locally and throughout the body.

Low-intensity ultraviolet is used to sterilize haircutting implements, implying that the energy associated with ultraviolet is deposited in a manner different from lower-frequency electromagnetic waves. (Actually this is true for all electromagnetic waves with frequencies greater than visible light.)

Flash photography is generally not allowed of precious artworks and colored prints because the UV radiation from the flash can cause photo-degradation in the artworks. Often artworks will have an extra-thick layer of glass in front of them, which is especially designed to absorb UV radiation.

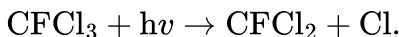
UV Light and the Ozone Layer

If all of the Sun's ultraviolet radiation reached the Earth's surface, there would be extremely grave effects on the biosphere from the severe cell damage it causes. However, the layer of ozone (O₃) in our upper atmosphere (10 to 50 km above the Earth) protects life by absorbing most of the dangerous UV radiation.

Unfortunately, today we are observing a depletion in ozone concentrations in the upper atmosphere. This depletion has led to the formation of an "ozone hole" in the upper atmosphere. The hole is more centered over the southern hemisphere, and changes with the seasons, being largest in the spring. This depletion is attributed to the breakdown of ozone molecules by refrigerant gases called chlorofluorocarbons (CFCs).

The UV radiation helps dissociate the CFC's, releasing highly reactive chlorine (Cl) atoms, which catalyze the destruction of the ozone layer. For example, the reaction of CFC₃ with a photon of light (hν) can be written as:

Equation:



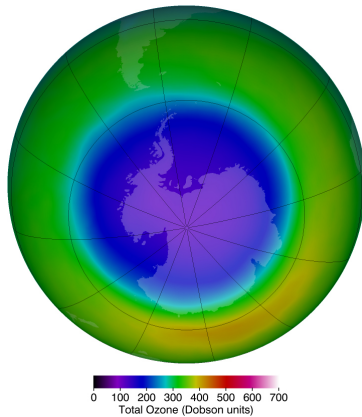
The Cl atom then catalyzes the breakdown of ozone as follows:

Equation:



A single chlorine atom could destroy ozone molecules for up to two years before being transported down to the surface. The CFCs are relatively stable and will contribute to ozone depletion for years to come. CFCs are found in refrigerants, air conditioning systems, foams, and aerosols.

International concern over this problem led to the establishment of the "Montreal Protocol" agreement (1987) to phase out CFC production in most countries. However, developing-country participation is needed if worldwide production and elimination of CFCs is to be achieved. Probably the largest contributor to CFC emissions today is India. But the protocol seems to be working, as there are signs of an ozone recovery. (See [\[link\]](#).)



This map of ozone concentration over Antarctica in October 2011 shows severe depletion suspected to be caused by CFCs.

Less dramatic but more general depletion has been observed over northern latitudes, suggesting the effect is global. With less ozone, more ultraviolet radiation from the Sun reaches the surface, causing more damage. (credit: NASA Ozone Watch)

Benefits of UV Light

Besides the adverse effects of ultraviolet radiation, there are also benefits of exposure in nature and uses in technology. Vitamin D production in the skin (epidermis) results from exposure to UVB radiation, generally from sunlight. A number of studies indicate lack of vitamin D can result in the development of a range of cancers (prostate, breast, colon), so a certain amount of UV exposure is helpful. Lack of vitamin D is also linked to osteoporosis. Exposures (with no sunscreen) of 10 minutes a day to arms, face, and legs might be sufficient to provide the accepted dietary level. However, in the winter time north of about 37° latitude, most UVB gets blocked by the atmosphere.

UV radiation is used in the treatment of infantile jaundice and in some skin conditions. It is also used in sterilizing workspaces and tools, and killing germs in a wide range of applications. It is

also used as an analytical tool to identify substances.

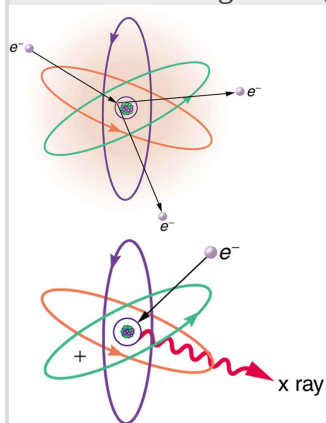
When exposed to ultraviolet, some substances, such as minerals, glow in characteristic visible wavelengths, a process called fluorescence. So-called black lights emit ultraviolet to cause posters and clothing to fluoresce in the visible. Ultraviolet is also used in special microscopes to detect details smaller than those observable with longer-wavelength visible-light microscopes.

Note:

Things Great and Small: A Submicroscopic View of X-Ray Production

X-rays can be created in a high-voltage discharge. They are emitted in the material struck by electrons in the discharge current. There are two mechanisms by which the electrons create X-rays.

The first method is illustrated in [\[link\]](#). An electron is accelerated in an evacuated tube by a high positive voltage. The electron strikes a metal plate (e.g., copper) and produces X-rays. Since this is a high-voltage discharge, the electron gains sufficient energy to ionize the atom.

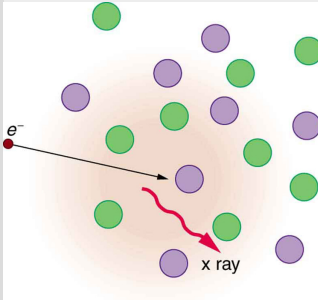


Artist's conception
of an electron
ionizing an atom
followed by the
recapture of an
electron and
emission of an X-
ray. An energetic
electron strikes an
atom and knocks an
electron out of one
of the orbits closest
to the nucleus.
Later, the atom
captures another
electron, and the
energy released by

its fall into a low orbit generates a high-energy EM wave called an X-ray.

In the case shown, an inner-shell electron (one in an orbit relatively close to and tightly bound to the nucleus) is ejected. A short time later, another electron is captured and falls into the orbit in a single great plunge. The energy released by this fall is given to an EM wave known as an X-ray. Since the orbits of the atom are unique to the type of atom, the energy of the X-ray is characteristic of the atom, hence the name characteristic X-ray.

The second method by which an energetic electron creates an X-ray when it strikes a material is illustrated in [\[link\]](#). The electron interacts with charges in the material as it penetrates. These collisions transfer kinetic energy from the electron to the electrons and atoms in the material.



Artist's conception of an electron being slowed by collisions in a material and emitting X-ray radiation. This energetic electron makes numerous collisions with electrons and atoms in a material it penetrates. An accelerated charge radiates EM waves, a second method by which X-rays are created.

A loss of kinetic energy implies an acceleration, in this case decreasing the electron's velocity. Whenever a charge is accelerated, it radiates EM waves. Given the high energy of the electron,

these EM waves can have high energy. We call them X-rays. Since the process is random, a broad spectrum of X-ray energy is emitted that is more characteristic of the electron energy than the type of material the electron encounters. Such EM radiation is called “bremsstrahlung” (German for “braking radiation”).

X-Rays

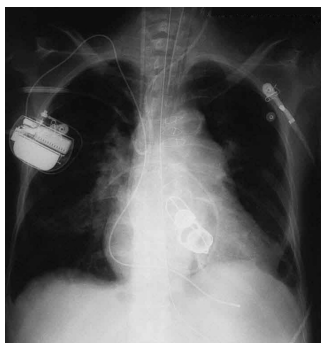
In the 1850s, scientists (such as Faraday) began experimenting with high-voltage electrical discharges in tubes filled with rarefied gases. It was later found that these discharges created an invisible, penetrating form of very high frequency electromagnetic radiation. This radiation was called an **X-ray**, because its identity and nature were unknown.

As described in [Things Great and Small](#), there are two methods by which X-rays are created—both are submicroscopic processes and can be caused by high-voltage discharges. While the low-frequency end of the X-ray range overlaps with the ultraviolet, X-rays extend to much higher frequencies (and energies).

X-rays have adverse effects on living cells similar to those of ultraviolet radiation, and they have the additional liability of being more penetrating, affecting more than the surface layers of cells. Cancer and genetic defects can be induced by exposure to X-rays. Because of their effect on rapidly dividing cells, X-rays can also be used to treat and even cure cancer.

The widest use of X-rays is for imaging objects that are opaque to visible light, such as the human body or aircraft parts. In humans, the risk of cell damage is weighed carefully against the benefit of the diagnostic information obtained. However, questions have risen in recent years as to accidental overexposure of some people during CT scans—a mistake at least in part due to poor monitoring of radiation dose.

The ability of X-rays to penetrate matter depends on density, and so an X-ray image can reveal very detailed density information. [\[link\]](#) shows an example of the simplest type of X-ray image, an X-ray shadow on film. The amount of information in a simple X-ray image is impressive, but more sophisticated techniques, such as CT scans, can reveal three-dimensional information with details smaller than a millimeter.



This shadow X-ray
image shows many
interesting features,
such as artificial
heart valves, a
pacemaker, and the
wires used to close
the sternum.
(credit: P. P. Urone)

The use of X-ray technology in medicine is called radiology—an established and relatively cheap tool in comparison to more sophisticated technologies. Consequently, X-rays are widely available and used extensively in medical diagnostics. During World War I, mobile X-ray units, advocated by Madame Marie Curie, were used to diagnose soldiers.

Because they can have wavelengths less than 0.01 nm, X-rays can be scattered (a process called X-ray diffraction) to detect the shape of molecules and the structure of crystals. X-ray diffraction was crucial to Crick, Watson, and Wilkins in the determination of the shape of the double-helix DNA molecule.

X-rays are also used as a precise tool for trace-metal analysis in X-ray induced fluorescence, in which the energy of the X-ray emissions are related to the specific types of elements and amounts of materials present.

Gamma Rays

Soon after nuclear radioactivity was first detected in 1896, it was found that at least three distinct types of radiation were being emitted. The most penetrating nuclear radiation was called a **gamma ray (γ ray)** (again a name given because its identity and character were unknown), and it was later found to be an extremely high frequency electromagnetic wave.

In fact, γ rays are any electromagnetic radiation emitted by a nucleus. This can be from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ -ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation.

Gamma rays have characteristics identical to X-rays of the same frequency—they differ only in source. At higher frequencies, γ rays are more penetrating and more damaging to living tissue. They have many of the same uses as X-rays, including cancer therapy. Gamma radiation from radioactive materials is used in nuclear medicine.

[\[link\]](#) shows a medical image based on γ rays. Food spoilage can be greatly inhibited by exposing it to large doses of γ radiation, thereby obliterating responsible microorganisms. Damage to food cells through irradiation occurs as well, and the long-term hazards of

consuming radiation-preserved food are unknown and controversial for some groups. Both X-ray and γ -ray technologies are also used in scanning luggage at airports.



This is an image of the γ rays emitted by nuclei in a compound that is concentrated in the bones and eliminated through the kidneys. Bone cancer is evidenced by nonuniform concentration in similar

structures.
For example,
some ribs are
darker than
others.
(credit: P. P.
Urone)

Detecting Electromagnetic Waves from Space

A final note on star gazing. The entire electromagnetic spectrum is used by researchers for investigating stars, space, and time. As noted earlier, Penzias and Wilson detected microwaves to identify the background radiation originating from the Big Bang. Radio telescopes such as the Arecibo Radio Telescope in Puerto Rico and Parkes Observatory in Australia were designed to detect radio waves.

Infrared telescopes need to have their detectors cooled by liquid nitrogen to be able to gather useful signals. Since infrared radiation is predominantly from thermal agitation, if the detectors were not cooled, the vibrations of the molecules in the antenna would be stronger than the signal being collected.

The most famous of these infrared sensitive telescopes is the James Clerk Maxwell Telescope in Hawaii. The earliest telescopes, developed in the seventeenth century, were optical telescopes, collecting visible light. Telescopes in the ultraviolet, X-ray, and γ -ray regions are placed outside the atmosphere on satellites orbiting the Earth.

The Hubble Space Telescope (launched in 1990) gathers ultraviolet radiation as well as visible light. In the X-ray region, there is the Chandra X-ray Observatory (launched in 1999), and in the γ -ray region, there is the new Fermi Gamma-ray Space Telescope (launched in 2008—taking the place of the Compton Gamma Ray Observatory, 1991–2000.).

Note:

PhET Explorations: Color Vision

Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

[Color
Vision](#)
[n](#)

Section Summary

- The relationship among the speed of propagation, wavelength, and frequency for any wave is given by $v_W = f\lambda$, so that for electromagnetic waves,

Equation:

$$c = f\lambda,$$

where f is the frequency, λ is the wavelength, and c is the speed of light.

- The electromagnetic spectrum is separated into many categories and subcategories, based on the frequency and wavelength, source, and uses of the electromagnetic waves.
- Any electromagnetic wave produced by currents in wires is classified as a radio wave, the lowest frequency electromagnetic waves. Radio waves are divided into many types, depending on their applications, ranging up to microwaves at their highest frequencies.
- Infrared radiation lies below visible light in frequency and is produced by thermal motion and the vibration and rotation of atoms and molecules. Infrared's lower frequencies overlap with the highest-frequency microwaves.
- Visible light is largely produced by electronic transitions in atoms and molecules, and is defined as being detectable by the human eye. Its colors vary with frequency, from red at the lowest to violet at the highest.
- Ultraviolet radiation starts with frequencies just above violet in the visible range and is produced primarily by electronic transitions in atoms and molecules.
- X-rays are created in high-voltage discharges and by electron bombardment of metal targets. Their lowest frequencies overlap the ultraviolet range but extend to much higher values, overlapping at the high end with gamma rays.
- Gamma rays are nuclear in origin and are defined to include the highest-frequency electromagnetic radiation of any type.

Conceptual Questions

Exercise:

Problem:

If you live in a region that has a particular TV station, you can sometimes pick up some of its audio portion on your FM radio receiver. Explain how this is possible. Does it imply that TV audio is broadcast as FM?

Exercise:

Problem:

Explain why people who have the lens of their eye removed because of cataracts are able to see low-frequency ultraviolet.

Exercise:

Problem:

How do fluorescent soap residues make clothing look “brighter and whiter” in outdoor light? Would this be effective in candlelight?

Exercise:

Problem: Give an example of resonance in the reception of electromagnetic waves.

Exercise:

Problem:

Illustrate that the size of details of an object that can be detected with electromagnetic waves is related to their wavelength, by comparing details observable with two different types (for example, radar and visible light or infrared and X-rays).

Exercise:

Problem: Why don’t buildings block radio waves as completely as they do visible light?

Exercise:

Problem:

Make a list of some everyday objects and decide whether they are transparent or opaque to each of the types of electromagnetic waves.

Exercise:

Problem:

Your friend says that more patterns and colors can be seen on the wings of birds if viewed in ultraviolet light. Would you agree with your friend? Explain your answer.

Exercise:

Problem:

The rate at which information can be transmitted on an electromagnetic wave is proportional to the frequency of the wave. Is this consistent with the fact that laser telephone transmission at visible frequencies carries far more conversations per optical fiber than conventional electronic transmission in a wire? What is the implication for ELF radio communication with submarines?

Exercise:

Problem: Give an example of energy carried by an electromagnetic wave.

Exercise:

Problem:

In an MRI scan, a higher magnetic field requires higher frequency radio waves to resonate with the nuclear type whose density and location is being imaged. What effect does going to a larger magnetic field have on the most efficient antenna to broadcast those radio waves? Does it favor a smaller or larger antenna?

Exercise:**Problem:**

Laser vision correction often uses an excimer laser that produces 193-nm electromagnetic radiation. This wavelength is extremely strongly absorbed by the cornea and ablates it in a manner that reshapes the cornea to correct vision defects. Explain how the strong absorption helps concentrate the energy in a thin layer and thus give greater accuracy in shaping the cornea. Also explain how this strong absorption limits damage to the lens and retina of the eye.

Problems & Exercises**Exercise:****Problem:**

(a) Two microwave frequencies are authorized for use in microwave ovens: 900 and 2560 MHz. Calculate the wavelength of each. (b) Which frequency would produce smaller hot spots in foods due to interference effects?

Solution:

(a) 33.3 cm (900 MHz) 11.7 cm (2560 MHz)

(b) The microwave oven with the smaller wavelength would produce smaller hot spots in foods, corresponding to the one with the frequency 2560 MHz.

Exercise:**Problem:**

(a) Calculate the range of wavelengths for AM radio given its frequency range is 540 to 1600 kHz. (b) Do the same for the FM frequency range of 88.0 to 108 MHz.

Exercise:**Problem:**

A radio station utilizes frequencies between commercial AM and FM. What is the frequency of a 11.12-m-wavelength channel?

Solution:

26.96 MHz

Exercise:

Problem:

Find the frequency range of visible light, given that it encompasses wavelengths from 380 to 760 nm.

Exercise:

Problem:

Combing your hair leads to excess electrons on the comb. How fast would you have to move the comb up and down to produce red light?

Solution:

$$5.0 \times 10^{14} \text{ Hz}$$

Exercise:

Problem:

Electromagnetic radiation having a $15.0 - \mu\text{m}$ wavelength is classified as infrared radiation. What is its frequency?

Exercise:

Problem:

Approximately what is the smallest detail observable with a microscope that uses ultraviolet light of frequency $1.20 \times 10^{15} \text{ Hz}$?

Solution:

Equation:

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^8 \text{ m/s}}{1.20 \times 10^{15} \text{ Hz}} = 2.50 \times 10^{-7} \text{ m}$$

Exercise:

Problem:

A radar used to detect the presence of aircraft receives a pulse that has reflected off an object $6 \times 10^{-5} \text{ s}$ after it was transmitted. What is the distance from the radar station to the reflecting object?

Exercise:

Problem:

Some radar systems detect the size and shape of objects such as aircraft and geological terrain. Approximately what is the smallest observable detail utilizing 500-MHz radar?

Solution:

0.600 m

Exercise:**Problem:**

Determine the amount of time it takes for X-rays of frequency 3×10^{18} Hz to travel (a) 1 mm and (b) 1 cm.

Exercise:**Problem:**

If you wish to detect details of the size of atoms (about 1×10^{-10} m) with electromagnetic radiation, it must have a wavelength of about this size. (a) What is its frequency? (b) What type of electromagnetic radiation might this be?

Solution:

$$(a) f = \frac{c}{\lambda} = \frac{3.00 \times 10^8 \text{ m/s}}{1 \times 10^{-10} \text{ m}} = 3 \times 10^{18} \text{ Hz}$$

(b) X-rays

Exercise:**Problem:**

If the Sun suddenly turned off, we would not know it until its light stopped coming. How long would that be, given that the Sun is 1.50×10^{11} m away?

Exercise:**Problem:**

Distances in space are often quoted in units of light years, the distance light travels in one year. (a) How many meters is a light year? (b) How many meters is it to Andromeda, the nearest large galaxy, given that it is 2.00×10^6 light years away? (c) The most distant galaxy yet discovered is 12.0×10^9 light years away. How far is this in meters?

Exercise:

Problem:

A certain 50.0-Hz AC power line radiates an electromagnetic wave having a maximum electric field strength of 13.0 kV/m. (a) What is the wavelength of this very low frequency electromagnetic wave? (b) What is its maximum magnetic field strength?

Solution:

(a) $6.00 \times 10^6 \text{ m}$

(b) $4.33 \times 10^{-5} \text{ T}$

Exercise:**Problem:**

During normal beating, the heart creates a maximum 4.00-mV potential across 0.300 m of a person's chest, creating a 1.00-Hz electromagnetic wave. (a) What is the maximum electric field strength created? (b) What is the corresponding maximum magnetic field strength in the electromagnetic wave? (c) What is the wavelength of the electromagnetic wave?

Exercise:**Problem:**

(a) The ideal size (most efficient) for a broadcast antenna with one end on the ground is one-fourth the wavelength ($\lambda/4$) of the electromagnetic radiation being sent out. If a new radio station has such an antenna that is 50.0 m high, what frequency does it broadcast most efficiently? Is this in the AM or FM band? (b) Discuss the analogy of the fundamental resonant mode of an air column closed at one end to the resonance of currents on an antenna that is one-fourth their wavelength.

Solution:

(a) $1.50 \times 10^6 \text{ Hz}$, AM band

(b) The resonance of currents on an antenna that is $1/4$ their wavelength is analogous to the fundamental resonant mode of an air column closed at one end, since the tube also has a length equal to $1/4$ the wavelength of the fundamental oscillation.

Exercise:**Problem:**

(a) What is the wavelength of 100-MHz radio waves used in an MRI unit? (b) If the frequencies are swept over a ± 1.00 range centered on 100 MHz, what is the range of wavelengths broadcast?

Exercise:

Problem:

(a) What is the frequency of the 193-nm ultraviolet radiation used in laser eye surgery? (b) Assuming the accuracy with which this EM radiation can ablate the cornea is directly proportional to wavelength, how much more accurate can this UV be than the shortest visible wavelength of light?

Solution:

(a) $1.55 \times 10^{15} \text{ Hz}$

(b) The shortest wavelength of visible light is 380 nm, so that

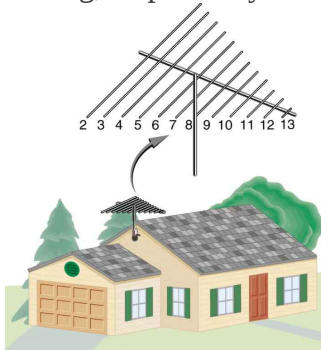
Equation:

$$\begin{aligned} \frac{\lambda_{\text{visible}}}{\lambda_{\text{UV}}} &= \frac{380 \text{ nm}}{193 \text{ nm}} \\ &= 1.97. \end{aligned}$$

In other words, the UV radiation is 97% more accurate than the shortest wavelength of visible light, or almost twice as accurate!

Exercise:**Problem:**

TV-reception antennas for VHF are constructed with cross wires supported at their centers, as shown in [\[link\]](#). The ideal length for the cross wires is one-half the wavelength to be received, with the more expensive antennas having one for each channel. Suppose you measure the lengths of the wires for particular channels and find them to be 1.94 and 0.753 m long, respectively. What are the frequencies for these channels?



A television reception antenna has cross wires of various lengths to most efficiently

receive different wavelengths.

Exercise:

Problem:

Conversations with astronauts on lunar walks had an echo that was used to estimate the distance to the Moon. The sound spoken by the person on Earth was transformed into a radio signal sent to the Moon, and transformed back into sound on a speaker inside the astronaut's space suit. This sound was picked up by the microphone in the space suit (intended for the astronaut's voice) and sent back to Earth as a radio echo of sorts. If the round-trip time was 2.60 s, what was the approximate distance to the Moon, neglecting any delays in the electronics?

Solution:

$$3.90 \times 10^8 \text{ m}$$

Exercise:

Problem:

Lunar astronauts placed a reflector on the Moon's surface, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. (a) To what accuracy in meters can the distance to the Moon be determined, if this time can be measured to 0.100 ns? (b) What percent accuracy is this, given the average distance to the Moon is $3.84 \times 10^8 \text{ m}$?

Exercise:

Problem:

Radar is used to determine distances to various objects by measuring the round-trip time for an echo from the object. (a) How far away is the planet Venus if the echo time is 1000 s? (b) What is the echo time for a car 75.0 m from a Highway Police radar unit? (c) How accurately (in nanoseconds) must you be able to measure the echo time to an airplane 12.0 km away to determine its distance within 10.0 m?

Solution:

(a) $1.50 \times 10^{11} \text{ m}$

(b) $0.500 \mu\text{s}$

(c) 66.7 ns

Exercise:

Problem: Integrated Concepts

- (a) Calculate the ratio of the highest to lowest frequencies of electromagnetic waves the eye can see, given the wavelength range of visible light is from 380 to 760 nm. (b) Compare this with the ratio of highest to lowest frequencies the ear can hear.

Exercise:**Problem: Integrated Concepts**

- (a) Calculate the rate in watts at which heat transfer through radiation occurs (almost entirely in the infrared) from 1.0 m^2 of the Earth's surface at night. Assume the emissivity is 0.90, the temperature of the Earth is 15°C , and that of outer space is 2.7 K. (b) Compare the intensity of this radiation with that coming to the Earth from the Sun during the day, which averages about 800 W/m^2 , only half of which is absorbed. (c) What is the maximum magnetic field strength in the outgoing radiation, assuming it is a continuous wave?

Solution:

- (a) $-3.5 \times 10^2 \text{ W/m}^2$
(b) 88%
(c) $1.7 \mu\text{T}$

Glossary

electromagnetic spectrum

the full range of wavelengths or frequencies of electromagnetic radiation

radio waves

electromagnetic waves with wavelengths in the range from 1 mm to 100 km; they are produced by currents in wires and circuits and by astronomical phenomena

microwaves

electromagnetic waves with wavelengths in the range from 1 mm to 1 m; they can be produced by currents in macroscopic circuits and devices

thermal agitation

the thermal motion of atoms and molecules in any object at a temperature above absolute zero, which causes them to emit and absorb radiation

radar

a common application of microwaves. Radar can determine the distance to objects as diverse as clouds and aircraft, as well as determine the speed of a car or the intensity of a

rainstorm

infrared radiation (IR)

a region of the electromagnetic spectrum with a frequency range that extends from just below the red region of the visible light spectrum up to the microwave region, or from $0.74\ \mu\text{m}$ to $300\ \mu\text{m}$

ultraviolet radiation (UV)

electromagnetic radiation in the range extending upward in frequency from violet light and overlapping with the lowest X-ray frequencies, with wavelengths from 400 nm down to about 10 nm

visible light

the narrow segment of the electromagnetic spectrum to which the normal human eye responds

amplitude modulation (AM)

a method for placing information on electromagnetic waves by modulating the amplitude of a carrier wave with an audio signal, resulting in a wave with constant frequency but varying amplitude

extremely low frequency (ELF)

electromagnetic radiation with wavelengths usually in the range of 0 to 300 Hz, but also about 1kHz

carrier wave

an electromagnetic wave that carries a signal by modulation of its amplitude or frequency

frequency modulation (FM)

a method of placing information on electromagnetic waves by modulating the frequency of a carrier wave with an audio signal, producing a wave of constant amplitude but varying frequency

TV

video and audio signals broadcast on electromagnetic waves

very high frequency (VHF)

TV channels utilizing frequencies in the two ranges of 54 to 88 MHz and 174 to 222 MHz

ultra-high frequency (UHF)

TV channels in an even higher frequency range than VHF, of 470 to 1000 MHz

X-ray

invisible, penetrating form of very high frequency electromagnetic radiation, overlapping both the ultraviolet range and the γ -ray range

gamma ray

(γ ray); extremely high frequency electromagnetic radiation emitted by the nucleus of an atom, either from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ -ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation

Energy in Electromagnetic Waves

- Explain how the energy and amplitude of an electromagnetic wave are related.
- Given its power output and the heating area, calculate the intensity of a microwave oven's electromagnetic field, as well as its peak electric and magnetic field strengths

Anyone who has used a microwave oven knows there is energy in **electromagnetic waves**. Sometimes this energy is obvious, such as in the warmth of the summer sun. Other times it is subtle, such as the unfelt energy of gamma rays, which can destroy living cells.

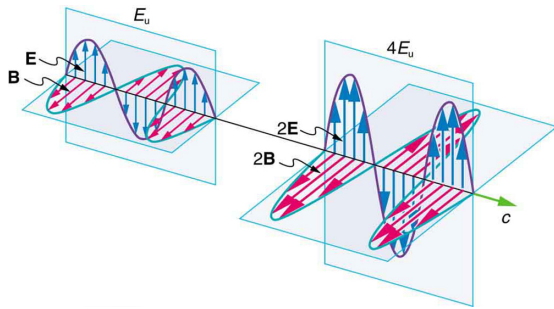
Electromagnetic waves can bring energy into a system by virtue of their **electric and magnetic fields**. These fields can exert forces and move charges in the system and, thus, do work on them. If the frequency of the electromagnetic wave is the same as the natural frequencies of the system (such as microwaves at the resonant frequency of water molecules), the transfer of energy is much more efficient.

Note:

Connections: Waves and Particles

The behavior of electromagnetic radiation clearly exhibits wave characteristics. But we shall find in later modules that at high frequencies, electromagnetic radiation also exhibits particle characteristics. These particle characteristics will be used to explain more of the properties of the electromagnetic spectrum and to introduce the formal study of modern physics.

Another startling discovery of modern physics is that particles, such as electrons and protons, exhibit wave characteristics. This simultaneous sharing of wave and particle properties for all submicroscopic entities is one of the great symmetries in nature.



Energy carried by a wave is proportional to its amplitude squared. With electromagnetic waves, larger E -fields and B -fields exert larger forces and can do more work.

But there is energy in an electromagnetic wave, whether it is absorbed or not. Once created, the fields carry energy away from a source. If absorbed, the field strengths are diminished and anything left travels on. Clearly, the larger the strength of the electric and magnetic fields, the more work they can do and the greater the energy the electromagnetic wave carries.

A wave's energy is proportional to its **amplitude** squared (E^2 or B^2). This is true for waves on guitar strings, for water waves, and for sound waves, where amplitude is proportional to pressure. In electromagnetic waves, the amplitude is the **maximum field strength** of the electric and magnetic fields. (See [\[link\]](#).)

Thus the energy carried and the **intensity** I of an electromagnetic wave is proportional to E^2 and B^2 . In fact, for a continuous sinusoidal electromagnetic wave, the average intensity I_{ave} is given by

Equation:

$$I_{\text{ave}} = \frac{c\epsilon_0 E_0^2}{2},$$

where c is the speed of light, ϵ_0 is the permittivity of free space, and E_0 is the maximum electric field strength; intensity, as always, is power per unit area (here in W/m^2).

The average intensity of an electromagnetic wave I_{ave} can also be expressed in terms of the magnetic field strength by using the relationship $B = E/c$, and the fact that $\epsilon_0 = 1/\mu_0 c^2$, where μ_0 is the permeability of free space. Algebraic manipulation produces the relationship

Equation:

$$I_{\text{ave}} = \frac{cB_0^2}{2\mu_0},$$

where B_0 is the maximum magnetic field strength.

One more expression for I_{ave} in terms of both electric and magnetic field strengths is useful. Substituting the fact that $c \cdot B_0 = E_0$, the previous expression becomes

Equation:

$$I_{\text{ave}} = \frac{E_0 B_0}{2\mu_0}.$$

Whichever of the three preceding equations is most convenient can be used, since they are really just different versions of the same principle: Energy in a wave is related to amplitude squared. Furthermore, since these equations are based on the assumption that the electromagnetic waves are sinusoidal, peak intensity is twice the average; that is, $I_0 = 2I_{\text{ave}}$.

Example:

Calculate Microwave Intensities and Fields

On its highest power setting, a certain microwave oven projects 1.00 kW of microwaves onto a 30.0 by 40.0 cm area. (a) What is the intensity in

W/m²? (b) Calculate the peak electric field strength E_0 in these waves.
(c) What is the peak magnetic field strength B_0 ?

Strategy

In part (a), we can find intensity from its definition as power per unit area. Once the intensity is known, we can use the equations below to find the field strengths asked for in parts (b) and (c).

Solution for (a)

Entering the given power into the definition of intensity, and noting the area is 0.300 by 0.400 m, yields

Equation:

$$I = \frac{P}{A} = \frac{1.00 \text{ kW}}{0.300 \text{ m} \times 0.400 \text{ m}}.$$

Here $I = I_{\text{ave}}$, so that

Equation:

$$I_{\text{ave}} = \frac{1000 \text{ W}}{0.120 \text{ m}^2} = 8.33 \times 10^3 \text{ W/m}^2.$$

Note that the peak intensity is twice the average:

Equation:

$$I_0 = 2I_{\text{ave}} = 1.67 \times 10^4 \text{ W/m}^2.$$

Solution for (b)

To find E_0 , we can rearrange the first equation given above for I_{ave} to give

Equation:

$$E_0 = \left(\frac{2I_{\text{ave}}}{c\epsilon_0} \right)^{1/2}.$$

Entering known values gives

Equation:

$$\begin{aligned}
 E_0 &= \sqrt{\frac{2(8.33 \times 10^3 \text{ W/m}^2)}{(3.00 \times 10^8 \text{ m/s})(8.85 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)}} \\
 &= 2.51 \times 10^3 \text{ V/m}.
 \end{aligned}$$

Solution for (c)

Perhaps the easiest way to find magnetic field strength, now that the electric field strength is known, is to use the relationship given by

Equation:

$$B_0 = \frac{E_0}{c}.$$

Entering known values gives

Equation:

$$\begin{aligned}
 B_0 &= \frac{2.51 \times 10^3 \text{ V/m}}{3.0 \times 10^8 \text{ m/s}} \\
 &= 8.35 \times 10^{-6} \text{ T}.
 \end{aligned}$$

Discussion

As before, a relatively strong electric field is accompanied by a relatively weak magnetic field in an electromagnetic wave, since $B = E/c$, and c is a large number.

Section Summary

- The energy carried by any wave is proportional to its amplitude squared. For electromagnetic waves, this means intensity can be expressed as

Equation:

$$I_{\text{ave}} = \frac{c\epsilon_0 E_0^2}{2},$$

where I_{ave} is the average intensity in W/m^2 , and E_0 is the maximum electric field strength of a continuous sinusoidal wave.

- This can also be expressed in terms of the maximum magnetic field strength B_0 as

Equation:

$$I_{\text{ave}} = \frac{cB_0^2}{2\mu_0}$$

and in terms of both electric and magnetic fields as

Equation:

$$I_{\text{ave}} = \frac{E_0 B_0}{2\mu_0}.$$

- The three expressions for I_{ave} are all equivalent.

Problems & Exercises

Exercise:

Problem:

What is the intensity of an electromagnetic wave with a peak electric field strength of 125 V/m?

Solution:

Equation:

$$\begin{aligned} I &= \frac{c\varepsilon_0 E_0^2}{2} \\ &= \frac{(3.00 \times 10^8 \text{ m/s})(8.85 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)(125 \text{ V/m})^2}{2} \\ &= 20.7 \text{ W/m}^2 \end{aligned}$$

Exercise:

Problem:

Find the intensity of an electromagnetic wave having a peak magnetic field strength of $4.00 \times 10^{-9} \text{ T}$.

Exercise:**Problem:**

Assume the helium-neon lasers commonly used in student physics laboratories have power outputs of 0.250 mW. (a) If such a laser beam is projected onto a circular spot 1.00 mm in diameter, what is its intensity? (b) Find the peak magnetic field strength. (c) Find the peak electric field strength.

Solution:

$$(a) \ I = \frac{P}{A} = \frac{P}{\pi r^2} = \frac{0.250 \times 10^{-3} \text{ W}}{\pi (0.500 \times 10^{-3} \text{ m})^2} = 318 \text{ W/m}^2$$

$$\begin{aligned} I_{\text{ave}} &= \frac{cB_0^2}{2\mu_0} \Rightarrow B_0 = \left(\frac{2\mu_0 I}{c} \right)^{1/2} \\ (b) \quad &= \left(\frac{2(4\pi \times 10^{-7} \text{ T}\cdot\text{m/A})(318.3 \text{ W/m}^2)}{3.00 \times 10^8 \text{ m/s}} \right)^{1/2} \\ &= 1.63 \times 10^{-6} \text{ T} \end{aligned}$$

$$\begin{aligned} (c) \quad E_0 &= cB_0 = (3.00 \times 10^8 \text{ m/s})(1.633 \times 10^{-6} \text{ T}) \\ &= 4.90 \times 10^2 \text{ V/m} \end{aligned}$$

Exercise:

Problem:

An AM radio transmitter broadcasts 50.0 kW of power uniformly in all directions. (a) Assuming all of the radio waves that strike the ground are completely absorbed, and that there is no absorption by the atmosphere or other objects, what is the intensity 30.0 km away? (Hint: Half the power will be spread over the area of a hemisphere.) (b) What is the maximum electric field strength at this distance?

Exercise:**Problem:**

Suppose the maximum safe intensity of microwaves for human exposure is taken to be 1.00 W/m^2 . (a) If a radar unit leaks 10.0 W of microwaves (other than those sent by its antenna) uniformly in all directions, how far away must you be to be exposed to an intensity considered to be safe? Assume that the power spreads uniformly over the area of a sphere with no complications from absorption or reflection. (b) What is the maximum electric field strength at the safe intensity? (Note that early radar units leaked more than modern ones do. This caused identifiable health problems, such as cataracts, for people who worked near them.)

Solution:

(a) 89.2 cm

(b) 27.4 V/m

Exercise:

Problem:

A 2.50-m-diameter university communications satellite dish receives TV signals that have a maximum electric field strength (for one channel) of $7.50\ \mu\text{V}/\text{m}$. (See [\[link\]](#).) (a) What is the intensity of this wave? (b) What is the power received by the antenna? (c) If the orbiting satellite broadcasts uniformly over an area of $1.50 \times 10^{13}\ \text{m}^2$ (a large fraction of North America), how much power does it radiate?



Satellite dishes receive TV signals sent from orbit. Although the signals are quite weak, the receiver can detect them by being tuned to resonate at their frequency.

Exercise:

Problem:

Lasers can be constructed that produce an extremely high intensity electromagnetic wave for a brief time—called pulsed lasers. They are used to ignite nuclear fusion, for example. Such a laser may produce an electromagnetic wave with a maximum electric field strength of $1.00 \times 10^{11} \text{ V/m}$ for a time of 1.00 ns. (a) What is the maximum magnetic field strength in the wave? (b) What is the intensity of the beam? (c) What energy does it deliver on a 1.00-mm^2 area?

Solution:

(a) 333 T

(b) $1.33 \times 10^{19} \text{ W/m}^2$

(c) 13.3 kJ

Exercise:**Problem:**

Show that for a continuous sinusoidal electromagnetic wave, the peak intensity is twice the average intensity ($I_0 = 2I_{\text{ave}}$), using either the fact that $E_0 = \sqrt{2}E_{\text{rms}}$, or $B_0 = \sqrt{2}B_{\text{rms}}$, where rms means average (actually root mean square, a type of average).

Exercise:**Problem:**

Suppose a source of electromagnetic waves radiates uniformly in all directions in empty space where there are no absorption or interference effects. (a) Show that the intensity is inversely proportional to r^2 , the distance from the source squared. (b) Show that the magnitudes of the electric and magnetic fields are inversely proportional to r .

Solution:

$$(a) I = \frac{P}{A} = \frac{P}{4\pi r^2} \propto \frac{1}{r^2}$$

$$(b) I \propto E_0^2, B_0^2 \Rightarrow E_0^2, B_0^2 \propto \frac{1}{r^2} \Rightarrow E_0, B_0 \propto \frac{1}{r}$$

Exercise:

Problem: Integrated Concepts

An LC circuit with a 5.00-pF capacitor oscillates in such a manner as to radiate at a wavelength of 3.30 m. (a) What is the resonant frequency? (b) What inductance is in series with the capacitor?

Exercise:

Problem: Integrated Concepts

What capacitance is needed in series with an $800\text{ } \mu\text{H}$ inductor to form a circuit that radiates a wavelength of 196 m?

Solution:

13.5 pF

Exercise:

Problem: Integrated Concepts

Police radar determines the speed of motor vehicles using the same Doppler-shift technique employed for ultrasound in medical diagnostics. Beats are produced by mixing the double Doppler-shifted echo with the original frequency. If 1.50×10^9 -Hz microwaves are used and a beat frequency of 150 Hz is produced, what is the speed of the vehicle? (Assume the same Doppler-shift formulas are valid with the speed of sound replaced by the speed of light.)

Exercise:

Problem: Integrated Concepts

Assume the mostly infrared radiation from a heat lamp acts like a continuous wave with wavelength $1.50\ \mu\text{m}$. (a) If the lamp's 200-W output is focused on a person's shoulder, over a circular area 25.0 cm in diameter, what is the intensity in W/m^2 ? (b) What is the peak electric field strength? (c) Find the peak magnetic field strength. (d) How long will it take to increase the temperature of the 4.00-kg shoulder by 2.00°C , assuming no other heat transfer and given that its specific heat is $3.47 \times 10^3\ \text{J}/\text{kg}\cdot^\circ\text{C}$?

Solution:

(a) $4.07\ \text{kW}/\text{m}^2$

(b) $1.75\ \text{kV}/\text{m}$

(c) $5.84\ \mu\text{T}$

(d) 2 min 19 s

Exercise:

Problem: Integrated Concepts

On its highest power setting, a microwave oven increases the temperature of 0.400 kg of spaghetti by 45.0°C in 120 s. (a) What was the rate of power absorption by the spaghetti, given that its specific heat is $3.76 \times 10^3\ \text{J}/\text{kg}\cdot^\circ\text{C}$? (b) Find the average intensity of the microwaves, given that they are absorbed over a circular area 20.0 cm in diameter. (c) What is the peak electric field strength of the microwave? (d) What is its peak magnetic field strength?

Exercise:

Problem: Integrated Concepts

Electromagnetic radiation from a 5.00-mW laser is concentrated on a 1.00-mm^2 area. (a) What is the intensity in W/m^2 ? (b) Suppose a 2.00-nC static charge is in the beam. What is the maximum electric

force it experiences? (c) If the static charge moves at 400 m/s, what maximum magnetic force can it feel?

Solution:

(a) $5.00 \times 10^3 \text{ W/m}^2$

(b) $3.88 \times 10^{-6} \text{ N}$

(c) $5.18 \times 10^{-12} \text{ N}$

Exercise:

Problem: Integrated Concepts

A 200-turn flat coil of wire 30.0 cm in diameter acts as an antenna for FM radio at a frequency of 100 MHz. The magnetic field of the incoming electromagnetic wave is perpendicular to the coil and has a maximum strength of $1.00 \times 10^{-12} \text{ T}$. (a) What power is incident on the coil? (b) What average emf is induced in the coil over one-fourth of a cycle? (c) If the radio receiver has an inductance of $2.50 \mu\text{H}$, what capacitance must it have to resonate at 100 MHz?

Exercise:

Problem: Integrated Concepts

If electric and magnetic field strengths vary sinusoidally in time, being zero at $t = 0$, then $E = E_0 \sin 2\pi ft$ and $B = B_0 \sin 2\pi ft$. Let $f = 1.00 \text{ GHz}$ here. (a) When are the field strengths first zero? (b) When do they reach their most negative value? (c) How much time is needed for them to complete one cycle?

Solution:

(a) $t = 0$

(b) $7.50 \times 10^{-10} \text{ s}$

(c) $1.00 \times 10^{-9} \text{ s}$

Exercise:

Problem: Unreasonable Results

A researcher measures the wavelength of a 1.20-GHz electromagnetic wave to be 0.500 m. (a) Calculate the speed at which this wave propagates. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: Unreasonable Results

The peak magnetic field strength in a residential microwave oven is $9.20 \times 10^{-5} \text{ T}$. (a) What is the intensity of the microwave? (b) What is unreasonable about this result? (c) What is wrong about the premise?

Solution:

(a) $1.01 \times 10^6 \text{ W/m}^2$

(b) Much too great for an oven.

(c) The assumed magnetic field is unreasonably large.

Exercise:

Problem: Unreasonable Results

An LC circuit containing a 2.00-H inductor oscillates at such a frequency that it radiates at a 1.00-m wavelength. (a) What is the capacitance of the circuit? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: Unreasonable Results

An LC circuit containing a 1.00-pF capacitor oscillates at such a frequency that it radiates at a 300-nm wavelength. (a) What is the inductance of the circuit? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $2.53 \times 10^{-20} \text{ H}$

(b) L is much too small.

(c) The wavelength is unreasonably small.

Exercise:

Problem: Create Your Own Problem

Consider electromagnetic fields produced by high voltage power lines. Construct a problem in which you calculate the intensity of this electromagnetic radiation in W/m^2 based on the measured magnetic field strength of the radiation in a home near the power lines. Assume these magnetic field strengths are known to average less than a μT . The intensity is small enough that it is difficult to imagine mechanisms for biological damage due to it. Discuss how much energy may be radiating from a section of power line several hundred meters long and compare this to the power likely to be carried by the lines. An idea of how much power this is can be obtained by calculating the approximate current responsible for μT fields at distances of tens of meters.

Exercise:

Problem: Create Your Own Problem

Consider the most recent generation of residential satellite dishes that are a little less than half a meter in diameter. Construct a problem in which you calculate the power received by the dish and the maximum electric field strength of the microwave signals for a single channel

received by the dish. Among the things to be considered are the power broadcast by the satellite and the area over which the power is spread, as well as the area of the receiving dish.

Glossary

maximum field strength

the maximum amplitude an electromagnetic wave can reach, representing the maximum amount of electric force and/or magnetic flux that the wave can exert

intensity

the power of an electric or magnetic field per unit area, for example, Watts per square meter

Introduction to Geometric Optics
class="introduction"

Geometric Optics

Light from this page or screen is formed into an image by the lens of your eye, much as the lens of the camera that made this photograph. Mirrors, like lenses, can also form images that in turn are captured by your eye.

Image
seen as a
result of
reflection
of light
on a
plane
smooth
surface.
(credit:
NASA
Goddard
Photo
and
Video,
via
Flickr)



Our lives are filled with light. Through vision, the most valued of our senses, light can evoke spiritual emotions, such as when we view a magnificent sunset or glimpse a rainbow breaking through the clouds. Light can also simply amuse us in a theater, or warn us to stop at an intersection. It has innumerable uses beyond vision. Light can carry telephone signals through glass fibers or cook a meal in a solar oven. Life itself could not exist without light's energy. From photosynthesis in plants to the sun warming a cold-blooded animal, its supply of energy is vital.



Double Rainbow over the bay

of Pocitos in Montevideo,
Uruguay. (credit: Madrax,
Wikimedia Commons)

We already know that visible light is the type of electromagnetic waves to which our eyes respond. That knowledge still leaves many questions regarding the nature of light and vision. What is color, and how do our eyes detect it? Why do diamonds sparkle? How does light travel? How do lenses and mirrors form images? These are but a few of the questions that are answered by the study of optics. Optics is the branch of physics that deals with the behavior of visible light and other electromagnetic waves. In particular, optics is concerned with the generation and propagation of light and its interaction with matter. What we have already learned about the generation of light in our study of heat transfer by radiation will be expanded upon in later topics, especially those on atomic physics. Now, we will concentrate on the propagation of light and its interaction with matter.

It is convenient to divide optics into two major parts based on the size of objects that light encounters. When light interacts with an object that is several times as large as the light's wavelength, its observable behavior is like that of a ray; it does not prominently display its wave characteristics. We call this part of optics "geometric optics." This chapter will concentrate on such situations. When light interacts with smaller objects, it has very prominent wave characteristics, such as constructive and destructive interference. [Wave Optics](#) will concentrate on such situations.

The Ray Aspect of Light

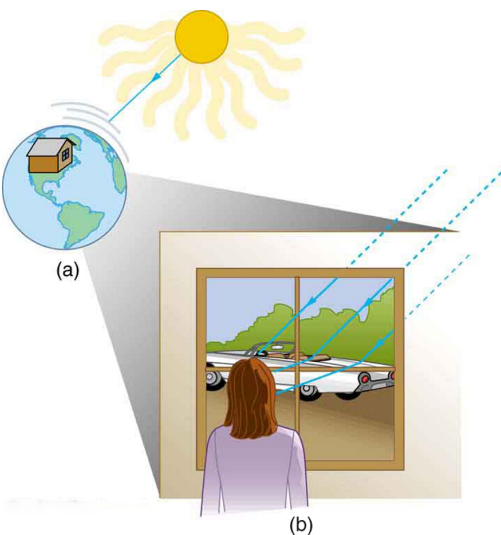
- List the ways by which light travels from a source to another location.

There are three ways in which light can travel from a source to another location. (See [\[link\]](#).) It can come directly from the source through empty space, such as from the Sun to Earth. Or light can travel through various media, such as air and glass, to the person. Light can also arrive after being reflected, such as by a mirror. In all of these cases, light is modeled as traveling in straight lines called rays. Light may change direction when it encounters objects (such as a mirror) or in passing from one material to another (such as in passing from air to glass), but it then continues in a straight line or as a ray. The word **ray** comes from mathematics and here means a straight line that originates at some point. It is acceptable to visualize light rays as laser rays (or even science fiction depictions of ray guns).

Note:

Ray

The word “ray” comes from mathematics and here means a straight line that originates at some point.



Three methods for light to travel from a source to another location. (a) Light reaches the upper atmosphere of Earth traveling through empty space directly from the source. (b) Light can reach a person in one of two ways. It can travel through media like air and glass. It can also reflect from an object like a mirror. In the situations shown here, light interacts with objects large enough that it travels in straight lines, like a ray.

Experiments, as well as our own experiences, show that when light interacts with objects several times as large as its wavelength, it travels in straight lines and acts like a ray. Its wave characteristics are not pronounced in such situations. Since the wavelength of light is less than a micron (a thousandth of a millimeter), it acts like a ray in the many common situations in which it encounters objects larger than a micron. For example, when light encounters anything we can observe with unaided eyes, such as a mirror, it acts like a ray, with only subtle wave characteristics. We will concentrate on the ray characteristics in this chapter.

Since light moves in straight lines, changing directions when it interacts with materials, it is described by geometry and simple trigonometry. This part of optics, where the ray aspect of light dominates, is therefore called **geometric optics**. There are two laws that govern how light changes direction when it interacts with matter. These are the law of reflection, for

situations in which light bounces off matter, and the law of refraction, for situations in which light passes through matter.

Note:

Geometric Optics

The part of optics dealing with the ray aspect of light is called geometric optics.

Section Summary

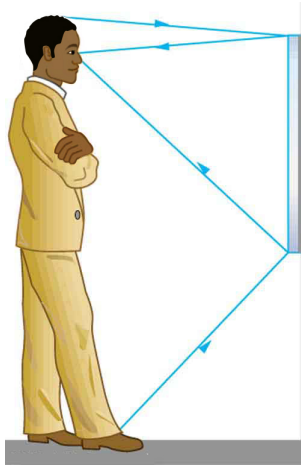
- A straight line that originates at some point is called a ray.
- The part of optics dealing with the ray aspect of light is called geometric optics.
- Light can travel in three ways from a source to another location: (1) directly from the source through empty space; (2) through various media; (3) after being reflected from a mirror.

Problems & Exercises

Exercise:

Problem:

Suppose a man stands in front of a mirror as shown in [\[link\]](#). His eyes are 1.65 m above the floor, and the top of his head is 0.13 m higher. Find the height above the floor of the top and bottom of the smallest mirror in which he can see both the top of his head and his feet. How is this distance related to the man's height?



A full-length mirror is one in which you can see all of yourself. It need not be as big as you, and its size is independent of your distance from it.

Solution:

Top 1.715 m from floor, bottom 0.825 m from floor. Height of mirror is 0.890 m, or precisely one-half the height of the person.

Glossary

ray

straight line that originates at some point

geometric optics

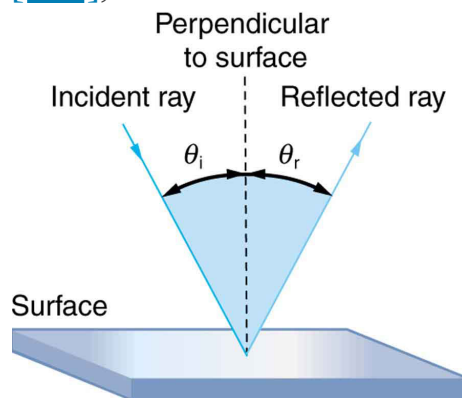
part of optics dealing with the ray aspect of light

The Law of Reflection

- Explain reflection of light from polished and rough surfaces.

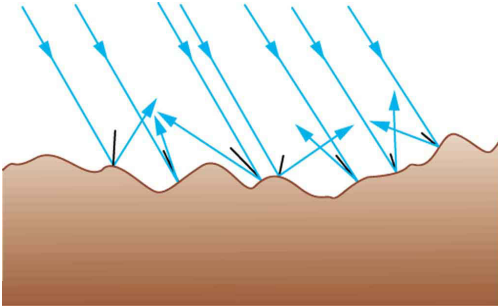
Whenever we look into a mirror, or squint at sunlight glinting from a lake, we are seeing a reflection. When you look at this page, too, you are seeing light reflected from it. Large telescopes use reflection to form an image of stars and other astronomical objects.

The law of reflection is illustrated in [\[link\]](#), which also shows how the angles are measured relative to the perpendicular to the surface at the point where the light ray strikes. We expect to see reflections from smooth surfaces, but [\[link\]](#) illustrates how a rough surface reflects light. Since the light strikes different parts of the surface at different angles, it is reflected in many different directions, or diffused. Diffused light is what allows us to see a sheet of paper from any angle, as illustrated in [\[link\]](#). Many objects, such as people, clothing, leaves, and walls, have rough surfaces and can be seen from all sides. A mirror, on the other hand, has a smooth surface (compared with the wavelength of light) and reflects light at specific angles, as illustrated in [\[link\]](#). When the moon reflects from a lake, as shown in [\[link\]](#), a combination of these effects takes place.

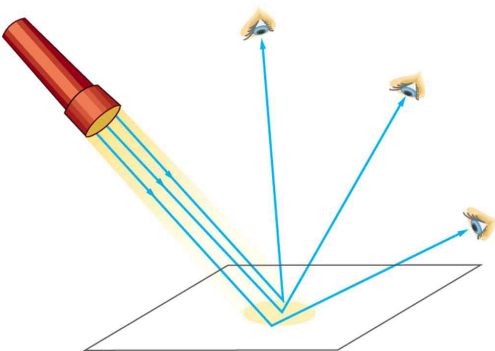


The law of reflection states that the angle of reflection equals the angle of incidence— $\theta_r = \theta_i$. The angles are measured relative to the perpendicular to

the surface at the point
where the ray strikes
the surface.

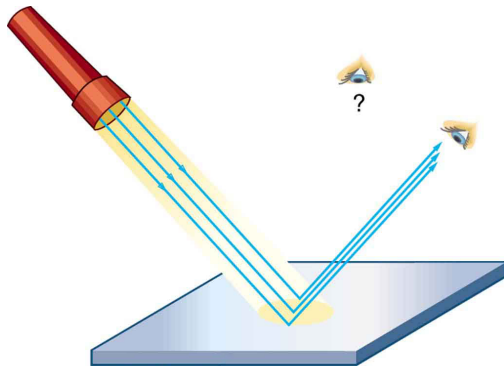


Light is diffused when it
reflects from a rough
surface. Here many
parallel rays are incident,
but they are reflected at
many different angles
since the surface is rough.

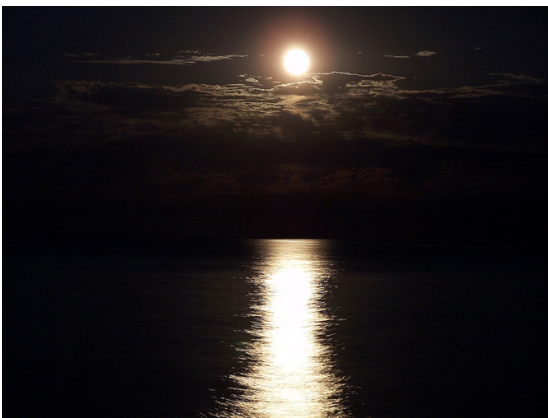


When a sheet of paper is
illuminated with many
parallel incident rays, it
can be seen at many
different angles, because

its surface is rough and
diffuses the light.



A mirror illuminated by
many parallel rays
reflects them in only one
direction, since its surface
is very smooth. Only the
observer at a particular
angle will see the
reflected light.



Moonlight is spread out
when it is reflected by the
lake, since the surface is
shiny but uneven. (credit:

Diego Torres Silvestre,
Flickr)

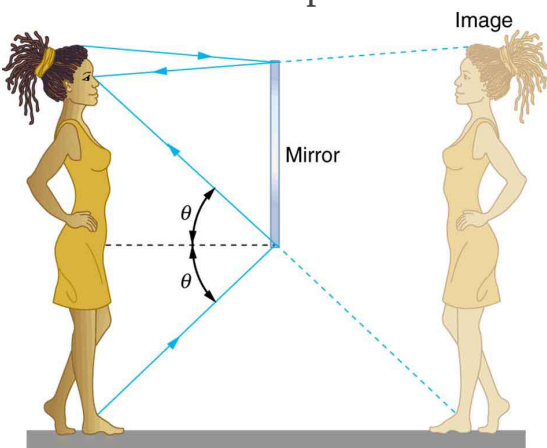
The law of reflection is very simple: The angle of reflection equals the angle of incidence.

Note:

The Law of Reflection

The angle of reflection equals the angle of incidence.

When we see ourselves in a mirror, it appears that our image is actually behind the mirror. This is illustrated in [\[link\]](#). We see the light coming from a direction determined by the law of reflection. The angles are such that our image is exactly the same distance behind the mirror as we stand away from the mirror. If the mirror is on the wall of a room, the images in it are all behind the mirror, which can make the room seem bigger. Although these mirror images make objects appear to be where they cannot be (like behind a solid wall), the images are not figments of our imagination. Mirror images can be photographed and videotaped by instruments and look just as they do with our eyes (optical instruments themselves). The precise manner in which images are formed by mirrors and lenses will be treated in later sections of this chapter.



Our image in a mirror is behind the mirror. The two rays shown are those that strike the mirror at just the correct angles to be reflected into the eyes of the person. The image appears to be in the direction the rays are coming from when they enter the eyes.

Note:**Take-Home Experiment: Law of Reflection**

Take a piece of paper and shine a flashlight at an angle at the paper, as shown in [\[link\]](#). Now shine the flashlight at a mirror at an angle. Do your observations confirm the predictions in [\[link\]](#) and [\[link\]](#)? Shine the flashlight on various surfaces and determine whether the reflected light is diffuse or not. You can choose a shiny metallic lid of a pot or your skin. Using the mirror and flashlight, can you confirm the law of reflection? You will need to draw lines on a piece of paper showing the incident and reflected rays. (This part works even better if you use a laser pencil.)

Section Summary

- The angle of reflection equals the angle of incidence.
- A mirror has a smooth surface and reflects light at specific angles.
- Light is diffused when it reflects from a rough surface.
- Mirror images can be photographed and videotaped by instruments.

Conceptual Questions

Exercise:

Problem:

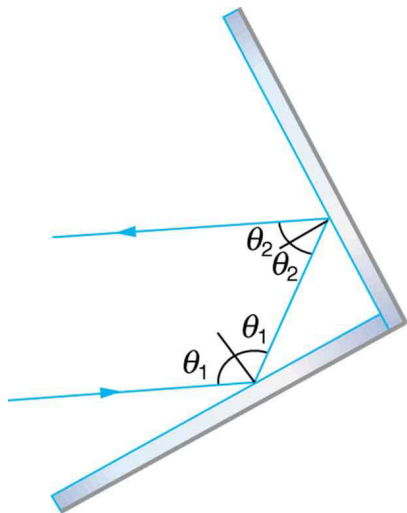
Using the law of reflection, explain how powder takes the shine off of a person's nose. What is the name of the optical effect?

Problems & Exercises

Exercise:

Problem:

Show that when light reflects from two mirrors that meet each other at a right angle, the outgoing ray is parallel to the incoming ray, as illustrated in the following figure.



A corner reflector sends the reflected ray back in a direction parallel to the incident ray, independent of incoming direction.

Exercise:

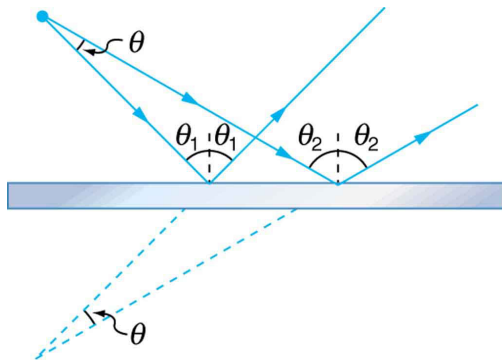
Problem:

Light shows staged with lasers use moving mirrors to swing beams and create colorful effects. Show that a light ray reflected from a mirror changes direction by 2θ when the mirror is rotated by an angle θ .

Exercise:

Problem:

A flat mirror is neither converging nor diverging. To prove this, consider two rays originating from the same point and diverging at an angle θ . Show that after striking a plane mirror, the angle between their directions remains θ .



A flat mirror neither converges nor diverges light rays. Two rays continue to diverge at the same angle after reflection.

Glossary

mirror

smooth surface that reflects light at specific angles, forming an image of the person or object in front of it

law of reflection

angle of reflection equals the angle of incidence

The Law of Refraction

- Determine the index of refraction, given the speed of light in a medium.

It is easy to notice some odd things when looking into a fish tank. For example, you may see the same fish appearing to be in two different places. (See [\[link\]](#).) This is because light coming from the fish to us changes direction when it leaves the tank, and in this case, it can travel two different paths to get to our eyes. The changing of a light ray's direction (loosely called bending) when it passes through variations in matter is called **refraction**. Refraction is responsible for a tremendous range of optical phenomena, from the action of lenses to voice transmission through optical fibers.

Note:

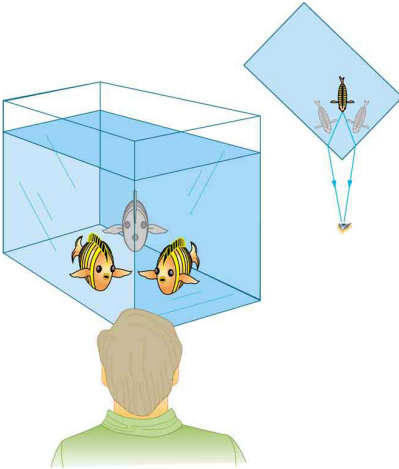
Refraction

The changing of a light ray's direction (loosely called bending) when it passes through variations in matter is called refraction.

Note:

Speed of Light

The speed of light c not only affects refraction, it is one of the central concepts of Einstein's theory of relativity. As the accuracy of the measurements of the speed of light were improved, c was found not to depend on the velocity of the source or the observer. However, the speed of light does vary in a precise manner with the material it traverses. These facts have far-reaching implications, as we will see in [Special Relativity](#). It makes connections between space and time and alters our expectations that all observers measure the same time for the same event, for example. The speed of light is so important that its value in a vacuum is one of the most fundamental constants in nature as well as being one of the four fundamental SI units.



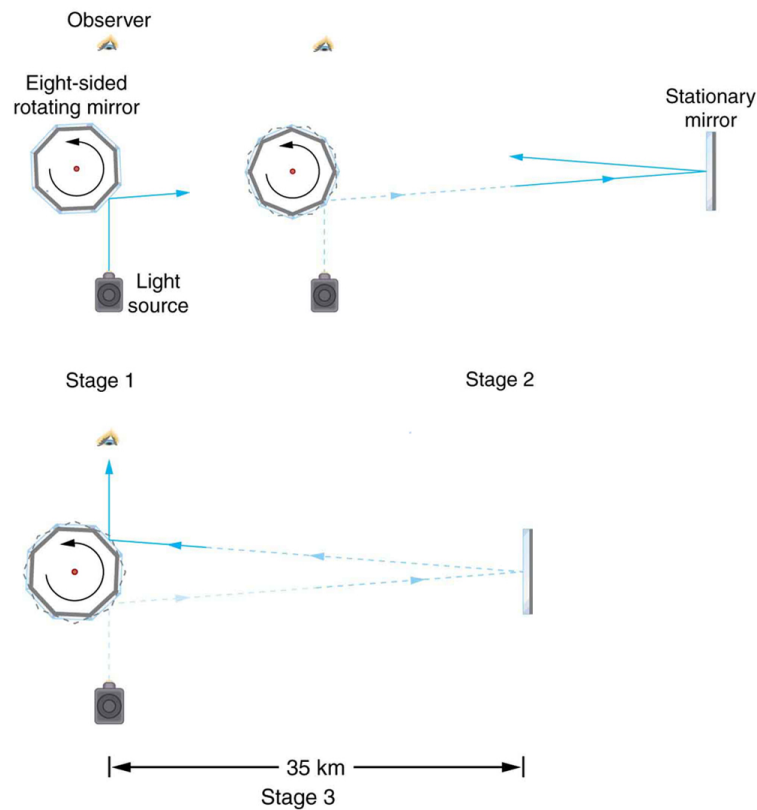
Looking at the fish tank as shown, we can see the same fish in two different locations, because light changes directions when it passes from water to air. In this case, the light can reach the observer by two different paths, and so the fish seems to be in two different places. This bending of light is called refraction and is responsible for many optical phenomena.

Why does light change direction when passing from one material (medium) to another? It is because light changes speed when going from one material

to another. So before we study the law of refraction, it is useful to discuss the speed of light and how it varies in different media.

The Speed of Light

Early attempts to measure the speed of light, such as those made by Galileo, determined that light moved extremely fast, perhaps instantaneously. The first real evidence that light traveled at a finite speed came from the Danish astronomer Ole Roemer in the late 17th century. Roemer had noted that the average orbital period of one of Jupiter's moons, as measured from Earth, varied depending on whether Earth was moving toward or away from Jupiter. He correctly concluded that the apparent change in period was due to the change in distance between Earth and Jupiter and the time it took light to travel this distance. From his 1676 data, a value of the speed of light was calculated to be 2.26×10^8 m/s (only 25% different from today's accepted value). In more recent times, physicists have measured the speed of light in numerous ways and with increasing accuracy. One particularly direct method, used in 1887 by the American physicist Albert Michelson (1852–1931), is illustrated in [\[link\]](#). Light reflected from a rotating set of mirrors was reflected from a stationary mirror 35 km away and returned to the rotating mirrors. The time for the light to travel can be determined by how fast the mirrors must rotate for the light to be returned to the observer's eye.



A schematic of early apparatus used by Michelson and others to determine the speed of light. As the mirrors rotate, the reflected ray is only briefly directed at the stationary mirror. The returning ray will be reflected into the observer's eye only if the next mirror has rotated into the correct position just as the ray returns. By measuring the correct rotation rate, the time for the round trip can be measured and the speed of light calculated. Michelson's calculated value of the speed of light was only 0.04% different from the value used today.

The speed of light is now known to great precision. In fact, the speed of light in a vacuum c is so important that it is accepted as one of the basic physical quantities and has the fixed value

Equation:

$$c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s},$$

where the approximate value of $3.00 \times 10^8 \text{ m/s}$ is used whenever three-digit accuracy is sufficient. The speed of light through matter is less than it is in a vacuum, because light interacts with atoms in a material. The speed of light depends strongly on the type of material, since its interaction with different atoms, crystal lattices, and other substructures varies. We define the **index of refraction** n of a material to be

Equation:

$$n = \frac{c}{v},$$

where v is the observed speed of light in the material. Since the speed of light is always less than c in matter and equals c only in a vacuum, the index of refraction is always greater than or equal to one.

Note:

Value of the Speed of Light

Equation:

$$c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$$

Note:

Index of Refraction

Equation:

$$n = \frac{c}{v}$$

That is, $n \geq 1$. [\[link\]](#) gives the indices of refraction for some representative substances. The values are listed for a particular wavelength of light, because they vary slightly with wavelength. (This can have important effects, such as colors produced by a prism.) Note that for gases, n is close to 1.0. This seems reasonable, since atoms in gases are widely separated and light travels at c in the vacuum between atoms. It is common to take $n = 1$ for gases unless great precision is needed. Although the speed of light v in a medium varies considerably from its value c in a vacuum, it is still a large speed.

Medium	n
<i>Gases at 0°C, 1 atm</i>	
Air	1.000293
Carbon dioxide	1.00045
Hydrogen	1.000139
Oxygen	1.000271
<i>Liquids at 20°C</i>	
Benzene	1.501
Carbon disulfide	1.628

Medium	<i>n</i>
Carbon tetrachloride	1.461
Ethanol	1.361
Glycerine	1.473
Water, fresh	1.333
<i>Solids at 20°C</i>	
Diamond	2.419
Fluorite	1.434
Glass, crown	1.52
Glass, flint	1.66
Ice at 20°C	1.309
Polystyrene	1.49
Plexiglas	1.51
Quartz, crystalline	1.544
Quartz, fused	1.458
Sodium chloride	1.544
Zircon	1.923

Index of Refraction in Various Media

Example:**Speed of Light in Matter**

Calculate the speed of light in zircon, a material used in jewelry to imitate diamond.

Strategy

The speed of light in a material, v , can be calculated from the index of refraction n of the material using the equation $n = c/v$.

Solution

The equation for index of refraction states that $n = c/v$. Rearranging this to determine v gives

Equation:

$$v = \frac{c}{n}.$$

The index of refraction for zircon is given as 1.923 in [\[link\]](#), and c is given in the equation for speed of light. Entering these values in the last expression gives

Equation:

$$\begin{aligned} v &= \frac{3.00 \times 10^8 \text{ m/s}}{1.923} \\ &= 1.56 \times 10^8 \text{ m/s.} \end{aligned}$$

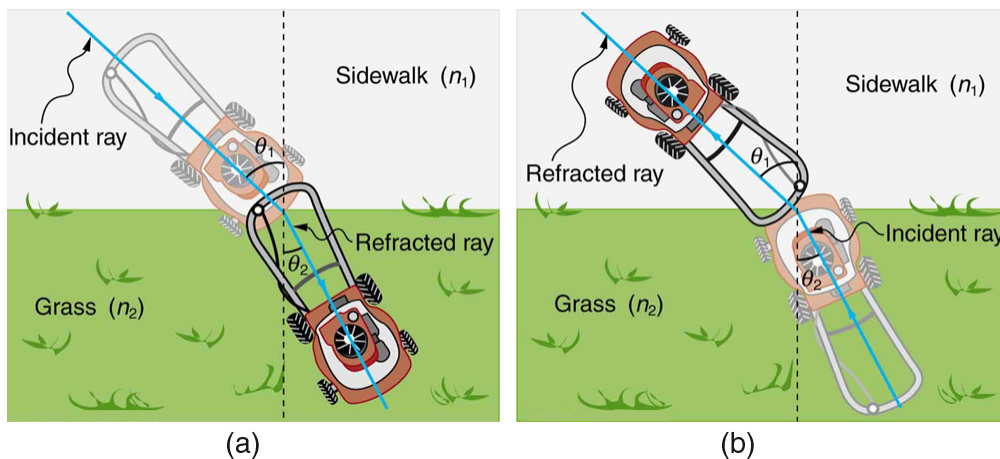
Discussion

This speed is slightly larger than half the speed of light in a vacuum and is still high compared with speeds we normally experience. The only substance listed in [\[link\]](#) that has a greater index of refraction than zircon is diamond. We shall see later that the large index of refraction for zircon makes it sparkle more than glass, but less than diamond.

Law of Refraction

[\[link\]](#) shows how a ray of light changes direction when it passes from one medium to another. As before, the angles are measured relative to a perpendicular to the surface at the point where the light ray crosses it.

(Some of the incident light will be reflected from the surface, but for now we will concentrate on the light that is transmitted.) The change in direction of the light ray depends on how the speed of light changes. The change in the speed of light is related to the indices of refraction of the media involved. In the situations shown in [\[link\]](#), medium 2 has a greater index of refraction than medium 1. This means that the speed of light is less in medium 2 than in medium 1. Note that as shown in [\[link\]](#)(a), the direction of the ray moves closer to the perpendicular when it slows down. Conversely, as shown in [\[link\]](#)(b), the direction of the ray moves away from the perpendicular when it speeds up. The path is exactly reversible. In both cases, you can imagine what happens by thinking about pushing a lawn mower from a footpath onto grass, and vice versa. Going from the footpath to grass, the front wheels are slowed and pulled to the side as shown. This is the same change in direction as for light when it goes from a fast medium to a slow one. When going from the grass to the footpath, the front wheels can move faster and the mower changes direction as shown. This, too, is the same change in direction as for light going from slow to fast.



The change in direction of a light ray depends on how the speed of light changes when it crosses from one medium to another. The speed of light is greater in medium 1 than in medium 2 in the situations shown here. (a) A ray of light moves closer to the perpendicular when it slows down. This is analogous to what happens when a lawn mower goes from a footpath to grass. (b) A ray of

light moves away from the perpendicular when it speeds up. This is analogous to what happens when a lawn mower goes from grass to footpath. The paths are exactly reversible.

The amount that a light ray changes its direction depends both on the incident angle and the amount that the speed changes. For a ray at a given incident angle, a large change in speed causes a large change in direction, and thus a large change in angle. The exact mathematical relationship is the **law of refraction**, or “Snell’s Law,” which is stated in equation form as **Equation:**

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

Here n_1 and n_2 are the indices of refraction for medium 1 and 2, and θ_1 and θ_2 are the angles between the rays and the perpendicular in medium 1 and 2, as shown in [\[link\]](#). The incoming ray is called the incident ray and the outgoing ray the refracted ray, and the associated angles the incident angle and the refracted angle. The law of refraction is also called Snell’s law after the Dutch mathematician Willebrord Snell (1591–1626), who discovered it in 1621. Snell’s experiments showed that the law of refraction was obeyed and that a characteristic index of refraction n could be assigned to a given medium. Snell was not aware that the speed of light varied in different media, but through experiments he was able to determine indices of refraction from the way light rays changed direction.

Note:

The Law of Refraction

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

Note:**Take-Home Experiment: A Broken Pencil**

A classic observation of refraction occurs when a pencil is placed in a glass half filled with water. Do this and observe the shape of the pencil when you look at the pencil sideways, that is, through air, glass, water. Explain your observations. Draw ray diagrams for the situation.

Example:**Determine the Index of Refraction from Refraction Data**

Find the index of refraction for medium 2 in [\[link\]](#)(a), assuming medium 1 is air and given the incident angle is 30.0° and the angle of refraction is 22.0° .

Strategy

The index of refraction for air is taken to be 1 in most cases (and up to four significant figures, it is 1.000). Thus $n_1 = 1.00$ here. From the given information, $\theta_1 = 30.0^\circ$ and $\theta_2 = 22.0^\circ$. With this information, the only unknown in Snell's law is n_2 , so that it can be used to find this unknown.

Solution

Snell's law is

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

Rearranging to isolate n_2 gives

Equation:

$$n_2 = n_1 \frac{\sin \theta_1}{\sin \theta_2}.$$

Entering known values,

Equation:

$$\begin{aligned} n_2 &= 1.00 \frac{\sin 30.0^\circ}{\sin 22.0^\circ} = \frac{0.500}{0.375} \\ &= 1.33. \end{aligned}$$

Discussion

This is the index of refraction for water, and Snell could have determined it by measuring the angles and performing this calculation. He would then have found 1.33 to be the appropriate index of refraction for water in all other situations, such as when a ray passes from water to glass. Today we can verify that the index of refraction is related to the speed of light in a medium by measuring that speed directly.

Example:**A Larger Change in Direction**

Suppose that in a situation like that in [\[link\]](#), light goes from air to diamond and that the incident angle is 30.0° . Calculate the angle of refraction θ_2 in the diamond.

Strategy

Again the index of refraction for air is taken to be $n_1 = 1.00$, and we are given $\theta_1 = 30.0^\circ$. We can look up the index of refraction for diamond in [\[link\]](#), finding $n_2 = 2.419$. The only unknown in Snell's law is θ_2 , which we wish to determine.

Solution

Solving Snell's law for $\sin \theta_2$ yields

Equation:

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1.$$

Entering known values,

Equation:

$$\sin \theta_2 = \frac{1.00}{2.419} \sin 30.0^\circ = (0.413)(0.500) = 0.207.$$

The angle is thus

Equation:

$$\theta_2 = \sin^{-1} 0.207 = 11.9^\circ.$$

Discussion

For the same 30° angle of incidence, the angle of refraction in diamond is significantly smaller than in water (11.9° rather than 22° —see the preceding example). This means there is a larger change in direction in diamond. The cause of a large change in direction is a large change in the index of refraction (or speed). In general, the larger the change in speed, the greater the effect on the direction of the ray.

Section Summary

- The changing of a light ray's direction when it passes through variations in matter is called refraction.
- The speed of light in vacuum
 $c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$.
- Index of refraction $n = \frac{c}{v}$, where v is the speed of light in the material, c is the speed of light in vacuum, and n is the index of refraction.
- Snell's law, the law of refraction, is stated in equation form as $n_1 \sin \theta_1 = n_2 \sin \theta_2$.

Conceptual Questions

Exercise:

Problem:

Diffusion by reflection from a rough surface is described in this chapter. Light can also be diffused by refraction. Describe how this occurs in a specific situation, such as light interacting with crushed ice.

Exercise:

Problem:

Why is the index of refraction always greater than or equal to 1?

Exercise:

Problem:

Does the fact that the light flash from lightning reaches you before its sound prove that the speed of light is extremely large or simply that it is greater than the speed of sound? Discuss how you could use this effect to get an estimate of the speed of light.

Exercise:**Problem:**

Will light change direction toward or away from the perpendicular when it goes from air to water? Water to glass? Glass to air?

Exercise:**Problem:**

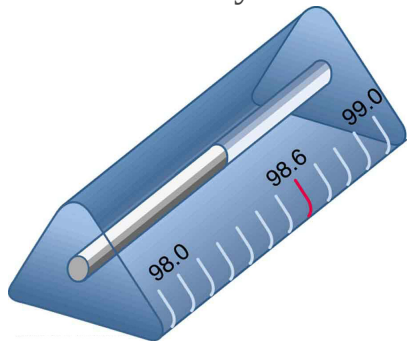
Explain why an object in water always appears to be at a depth shallower than it actually is? Why do people sometimes sustain neck and spinal injuries when diving into unfamiliar ponds or waters?

Exercise:**Problem:**

Explain why a person's legs appear very short when wading in a pool. Justify your explanation with a ray diagram showing the path of rays from the feet to the eye of an observer who is out of the water.

Exercise:

Problem: Why is the front surface of a thermometer curved as shown?



The curved surface
of the thermometer
serves a purpose.

Exercise:

Problem:

Suppose light were incident from air onto a material that had a negative index of refraction, say -1.3 ; where does the refracted light ray go?

Problems & Exercises

Exercise:

Problem: What is the speed of light in water? In glycerine?

Solution:

2.25×10^8 m/s in water

2.04×10^8 m/s in glycerine

Exercise:

Problem: What is the speed of light in air? In crown glass?

Exercise:

Problem:

Calculate the index of refraction for a medium in which the speed of light is 2.012×10^8 m/s, and identify the most likely substance based on [\[link\]](#).

Solution:

1.490, polystyrene

Exercise:**Problem:**

In what substance in [\[link\]](#) is the speed of light 2.290×10^8 m/s?

Exercise:**Problem:**

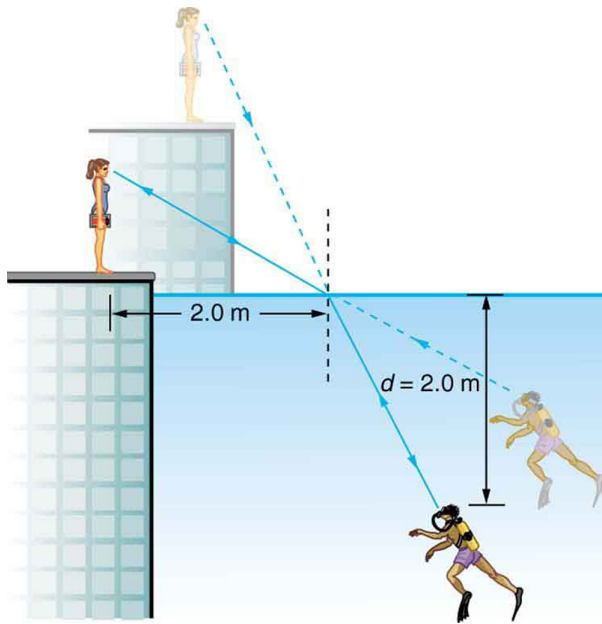
There was a major collision of an asteroid with the Moon in medieval times. It was described by monks at Canterbury Cathedral in England as a red glow on and around the Moon. How long after the asteroid hit the Moon, which is 3.84×10^5 km away, would the light first arrive on Earth?

Solution:

1.28 s

Exercise:**Problem:**

A scuba diver training in a pool looks at his instructor as shown in [\[link\]](#). What angle does the ray from the instructor's face make with the perpendicular to the water at the point where the ray enters? The angle between the ray in the water and the perpendicular to the water is 25.0° .



A scuba diver in a pool and his trainer look at each other.

Exercise:

Problem:

Components of some computers communicate with each other through optical fibers having an index of refraction $n = 1.55$. What time in nanoseconds is required for a signal to travel 0.200 m through such a fiber?

Solution:

1.03 ns

Exercise:

Problem:

(a) Given that the angle between the ray in the water and the perpendicular to the water is 25.0° , and using information in [\[link\]](#), find the height of the instructor's head above the water, noting that you will first have to calculate the angle of incidence. (b) Find the apparent depth of the diver's head below water as seen by the instructor.

Exercise:**Problem:**

Suppose you have an unknown clear substance immersed in water, and you wish to identify it by finding its index of refraction. You arrange to have a beam of light enter it at an angle of 45.0° , and you observe the angle of refraction to be 40.3° . What is the index of refraction of the substance and its likely identity?

Solution:

$n = 1.46$, fused quartz

Exercise:**Problem:**

On the Moon's surface, lunar astronauts placed a corner reflector, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. What percent correction is needed to account for the delay in time due to the slowing of light in Earth's atmosphere? Assume the distance to the Moon is precisely 3.84×10^8 m, and Earth's atmosphere (which varies in density with altitude) is equivalent to a layer 30.0 km thick with a constant index of refraction $n = 1.000293$.

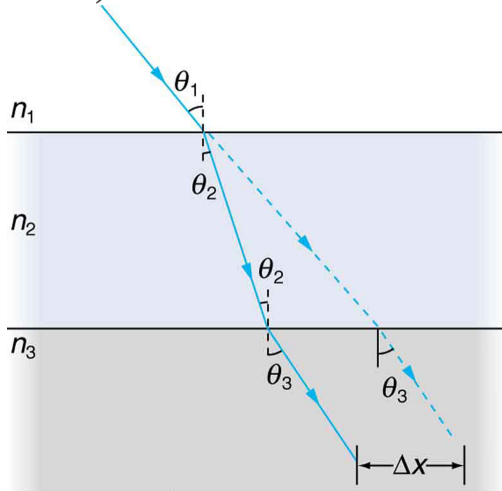
Exercise:

Problem:

Suppose [\[link\]](#) represents a ray of light going from air through crown glass into water, such as going into a fish tank. Calculate the amount the ray is displaced by the glass (Δx), given that the incident angle is 40.0° and the glass is 1.00 cm thick.

Exercise:**Problem:**

[\[link\]](#) shows a ray of light passing from one medium into a second and then a third. Show that θ_3 is the same as it would be if the second medium were not present (provided total internal reflection does not occur).



A ray of light passes from one medium to a third by traveling through a second. The final direction is the same as if the second medium were not present, but the ray is displaced by Δx (shown exaggerated).

Exercise:

Problem: Unreasonable Results

Suppose light travels from water to another substance, with an angle of incidence of 10.0° and an angle of refraction of 14.9° . (a) What is the index of refraction of the other substance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 0.898

(b) Can't have $n < 1.00$ since this would imply a speed greater than c .

(c) Refracted angle is too big relative to the angle of incidence.

Exercise:

Problem: Construct Your Own Problem

Consider sunlight entering the Earth's atmosphere at sunrise and sunset—that is, at a 90° incident angle. Taking the boundary between nearly empty space and the atmosphere to be sudden, calculate the angle of refraction for sunlight. This lengthens the time the Sun appears to be above the horizon, both at sunrise and sunset. Now construct a problem in which you determine the angle of refraction for different models of the atmosphere, such as various layers of varying density. Your instructor may wish to guide you on the level of complexity to consider and on how the index of refraction varies with air density.

Exercise:

Problem: Unreasonable Results

Light traveling from water to a gemstone strikes the surface at an angle of 80.0° and has an angle of refraction of 15.2° . (a) What is the speed

of light in the gemstone? (b) What is unreasonable about this result?
(c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $\frac{c}{5.00}$

(b) Speed of light too slow, since index is much greater than that of diamond.

(c) Angle of refraction is unreasonable relative to the angle of incidence.

Glossary

refraction

changing of a light ray's direction when it passes through variations in matter

index of refraction

for a material, the ratio of the speed of light in vacuum to that in the material

Total Internal Reflection

- Explain the phenomenon of total internal reflection.
- Describe the workings and uses of fiber optics.
- Analyze the reason for the sparkle of diamonds.

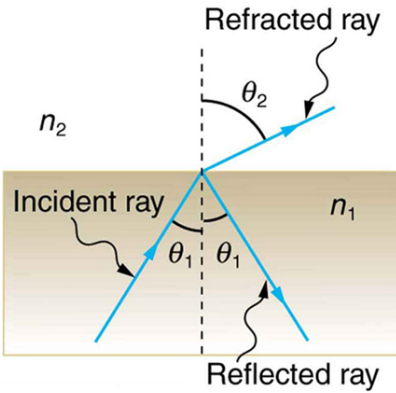
A good-quality mirror may reflect more than 90% of the light that falls on it, absorbing the rest. But it would be useful to have a mirror that reflects all of the light that falls on it. Interestingly, we can produce *total reflection* using an aspect of *refraction*.

Consider what happens when a ray of light strikes the surface between two materials, such as is shown in [\[link\]](#)(a). Part of the light crosses the boundary and is refracted; the rest is reflected. If, as shown in the figure, the index of refraction for the second medium is less than for the first, the ray bends away from the perpendicular. (Since $n_1 > n_2$, the angle of refraction is greater than the angle of incidence—that is, $\theta_2 > \theta_1$.) Now imagine what happens as the incident angle is increased. This causes θ_2 to increase also. The largest the angle of refraction θ_2 can be is 90° , as shown in [\[link\]](#)(b). The **critical angle** θ_c for a combination of materials is defined to be the incident angle θ_1 that produces an angle of refraction of 90° . That is, θ_c is the incident angle for which $\theta_2 = 90^\circ$. If the incident angle θ_1 is greater than the critical angle, as shown in [\[link\]](#)(c), then all of the light is reflected back into medium 1, a condition called **total internal reflection**.

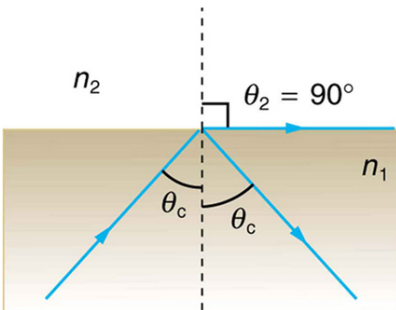
Note:

Critical Angle

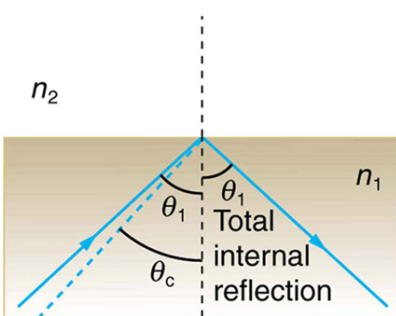
The incident angle θ_1 that produces an angle of refraction of 90° is called the critical angle, θ_c .



(a)



(b)



(c)

(a) A ray of light crosses a boundary where the speed of light increases and the index of refraction decreases. That is, $n_2 < n_1$. The ray bends away from the perpendicular.

(b) The critical

angle θ_c is the one for which the angle of refraction is 90° . (c)

Total internal reflection occurs when the incident angle is greater than the critical angle.

Snell's law states the relationship between angles and indices of refraction. It is given by

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

When the incident angle equals the critical angle ($\theta_1 = \theta_c$), the angle of refraction is 90° ($\theta_2 = 90^\circ$). Noting that $\sin 90^\circ = 1$, Snell's law in this case becomes

Equation:

$$n_1 \sin \theta_1 = n_2.$$

The critical angle θ_c for a given combination of materials is thus

Equation:

$$\theta_c = \sin^{-1}(n_2/n_1) \text{ for } n_1 > n_2.$$

Total internal reflection occurs for any incident angle greater than the critical angle θ_c , and it can only occur when the second medium has an index of refraction less than the first. Note the above equation is written for a light ray that travels in medium 1 and reflects from medium 2, as shown in the figure.

Example:**How Big is the Critical Angle Here?**

What is the critical angle for light traveling in a polystyrene (a type of plastic) pipe surrounded by air?

Strategy

The index of refraction for polystyrene is found to be 1.49 in [\[link\]](#), and the index of refraction of air can be taken to be 1.00, as before. Thus, the condition that the second medium (air) has an index of refraction less than the first (plastic) is satisfied, and the equation $\theta_c = \sin^{-1}(n_2/n_1)$ can be used to find the critical angle θ_c . Here, then, $n_2 = 1.00$ and $n_1 = 1.49$.

Solution

The critical angle is given by

Equation:

$$\theta_c = \sin^{-1}(n_2/n_1).$$

Substituting the identified values gives

Equation:

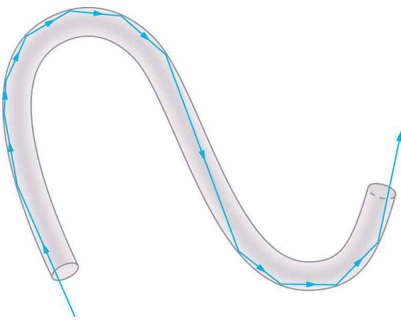
$$\theta_c = \sin^{-1}(1.00/1.49) = \sin^{-1}(0.671) \\ 42.2^\circ.$$

Discussion

This means that any ray of light inside the plastic that strikes the surface at an angle greater than 42.2° will be totally reflected. This will make the inside surface of the clear plastic a perfect mirror for such rays without any need for the silvering used on common mirrors. Different combinations of materials have different critical angles, but any combination with $n_1 > n_2$ can produce total internal reflection. The same calculation as made here shows that the critical angle for a ray going from water to air is 48.6° , while that from diamond to air is 24.4° , and that from flint glass to crown glass is 66.3° . There is no total reflection for rays going in the other direction—for example, from air to water—since the condition that the second medium must have a smaller index of refraction is not satisfied. A number of interesting applications of total internal reflection follow.

Fiber Optics: Endoscopes to Telephones

Fiber optics is one application of total internal reflection that is in wide use. In communications, it is used to transmit telephone, internet, and cable TV signals. **Fiber optics** employs the transmission of light down fibers of plastic or glass. Because the fibers are thin, light entering one is likely to strike the inside surface at an angle greater than the critical angle and, thus, be totally reflected (See [\[link\]](#).) The index of refraction outside the fiber must be smaller than inside, a condition that is easily satisfied by coating the outside of the fiber with a material having an appropriate refractive index. In fact, most fibers have a varying refractive index to allow more light to be guided along the fiber through total internal reflection. Rays are reflected around corners as shown, making the fibers into tiny light pipes.

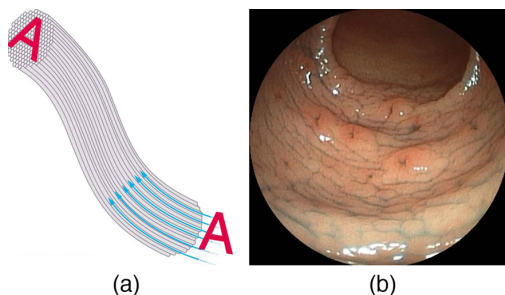


Light entering a thin fiber may strike the inside surface at large or grazing angles and is completely reflected if these angles exceed the critical angle. Such rays continue down the fiber, even following it around corners, since the angles of reflection

and incidence
remain large.

Bundles of fibers can be used to transmit an image without a lens, as illustrated in [\[link\]](#). The output of a device called an **endoscope** is shown in [\[link\]](#)(b). Endoscopes are used to explore the body through various orifices or minor incisions. Light is transmitted down one fiber bundle to illuminate internal parts, and the reflected light is transmitted back out through another to be observed. Surgery can be performed, such as arthroscopic surgery on the knee joint, employing cutting tools attached to and observed with the endoscope. Samples can also be obtained, such as by lassoing an intestinal polyp for external examination.

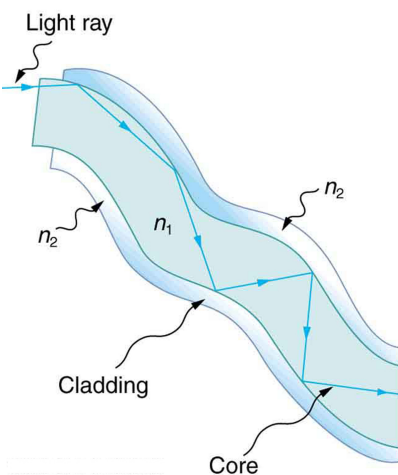
Fiber optics has revolutionized surgical techniques and observations within the body. There are a host of medical diagnostic and therapeutic uses. The flexibility of the fiber optic bundle allows it to navigate around difficult and small regions in the body, such as the intestines, the heart, blood vessels, and joints. Transmission of an intense laser beam to burn away obstructing plaques in major arteries as well as delivering light to activate chemotherapy drugs are becoming commonplace. Optical fibers have in fact enabled microsurgery and remote surgery where the incisions are small and the surgeon's fingers do not need to touch the diseased tissue.



(a) An image is
transmitted by a bundle of
fibers that have fixed

neighbors. (b) An endoscope is used to probe the body, both transmitting light to the interior and returning an image such as the one shown. (credit: Med_Chaos, Wikimedia Commons)

Fibers in bundles are surrounded by a cladding material that has a lower index of refraction than the core. (See [\[link\]](#).) The cladding prevents light from being transmitted between fibers in a bundle. Without cladding, light could pass between fibers in contact, since their indices of refraction are identical. Since no light gets into the cladding (there is total internal reflection back into the core), none can be transmitted between clad fibers that are in contact with one another. The cladding prevents light from escaping out of the fiber; instead most of the light is propagated along the length of the fiber, minimizing the loss of signal and ensuring that a quality image is formed at the other end. The cladding and an additional protective layer make optical fibers flexible and durable.



Fibers in bundles
are clad by a
material that has a
lower index of
refraction than the
core to ensure total
internal reflection,
even when fibers
are in contact with
one another. This
shows a single fiber
with its cladding.

Note:

Cladding

The cladding prevents light from being transmitted between fibers in a bundle.

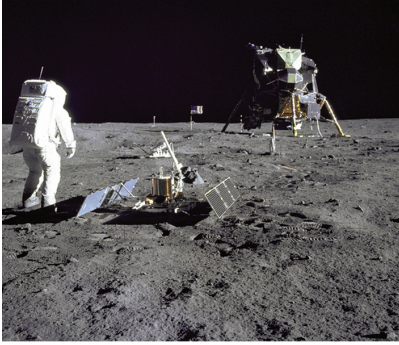
Special tiny lenses that can be attached to the ends of bundles of fibers are being designed and fabricated. Light emerging from a fiber bundle can be focused and a tiny spot can be imaged. In some cases the spot can be scanned, allowing quality imaging of a region inside the body. Special minute optical filters inserted at the end of the fiber bundle have the capacity to image tens of microns below the surface without cutting the surface—non-intrusive diagnostics. This is particularly useful for determining the extent of cancers in the stomach and bowel.

Most telephone conversations and Internet communications are now carried by laser signals along optical fibers. Extensive optical fiber cables have been placed on the ocean floor and underground to enable optical communications. Optical fiber communication systems offer several advantages over electrical (copper) based systems, particularly for long

distances. The fibers can be made so transparent that light can travel many kilometers before it becomes dim enough to require amplification—much superior to copper conductors. This property of optical fibers is called *low loss*. Lasers emit light with characteristics that allow far more conversations in one fiber than are possible with electric signals on a single conductor. This property of optical fibers is called *high bandwidth*. Optical signals in one fiber do not produce undesirable effects in other adjacent fibers. This property of optical fibers is called *reduced crosstalk*. We shall explore the unique characteristics of laser radiation in a later chapter.

Corner Reflectors and Diamonds

A light ray that strikes an object consisting of two mutually perpendicular reflecting surfaces is reflected back exactly parallel to the direction from which it came. This is true whenever the reflecting surfaces are perpendicular, and it is independent of the angle of incidence. Such an object, shown in [\[link\]](#), is called a **corner reflector**, since the light bounces from its inside corner. Many inexpensive reflector buttons on bicycles, cars, and warning signs have corner reflectors designed to return light in the direction from which it originated. It was more expensive for astronauts to place one on the moon. Laser signals can be bounced from that corner reflector to measure the gradually increasing distance to the moon with great precision.



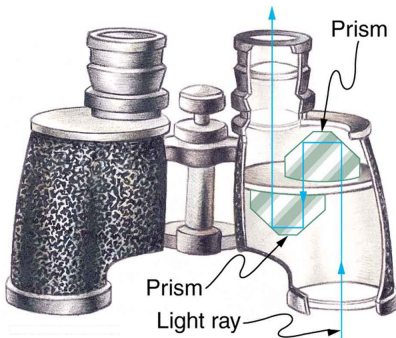
(a)



(b)

(a) Astronauts placed a corner reflector on the moon to measure its gradually increasing orbital distance. (credit: NASA) (b) The bright spots on these bicycle safety reflectors are reflections of the flash of the camera that took this picture on a dark night. (credit: Julo, Wikimedia Commons)

Corner reflectors are perfectly efficient when the conditions for total internal reflection are satisfied. With common materials, it is easy to obtain a critical angle that is less than 45° . One use of these perfect mirrors is in binoculars, as shown in [\[link\]](#). Another use is in periscopes found in submarines.

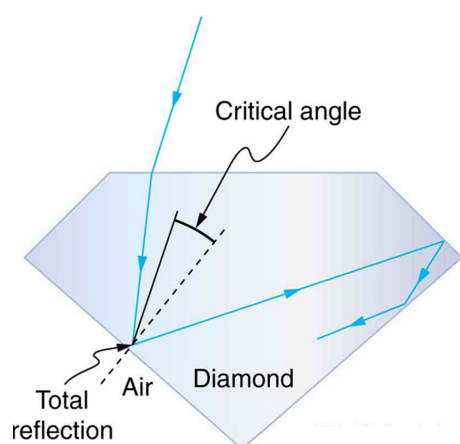


These binoculars employ corner reflectors with total internal reflection to get light to the observer's eyes.

The Sparkle of Diamonds

Total internal reflection, coupled with a large index of refraction, explains why diamonds sparkle more than other materials. The critical angle for a diamond-to-air surface is only 24.4° , and so when light enters a diamond, it has trouble getting back out. (See [\[link\]](#).) Although light freely enters the diamond, it can exit only if it makes an angle less than 24.4° . Facets on diamonds are specifically intended to make this unlikely, so that the light can exit only in certain places. Good diamonds are very clear, so that the light makes many internal reflections and is concentrated at the few places it can exit—hence the sparkle. (Zircon is a natural gemstone that has an exceptionally large index of refraction, but not as large as diamond, so it is

not as highly prized. Cubic zirconia is manufactured and has an even higher index of refraction (≈ 2.17), but still less than that of diamond.) The colors you see emerging from a sparkling diamond are not due to the diamond's color, which is usually nearly colorless. Those colors result from dispersion, the topic of [Dispersion: The Rainbow and Prisms](#). Colored diamonds get their color from structural defects of the crystal lattice and the inclusion of minute quantities of graphite and other materials. The Argyle Mine in Western Australia produces around 90% of the world's pink, red, champagne, and cognac diamonds, while around 50% of the world's clear diamonds come from central and southern Africa.



Light cannot easily escape a diamond, because its critical angle with air is so small. Most reflections are total, and the facets are placed so that light can exit only in particular ways—thus concentrating the light and making the diamond sparkle.

Note:**PhET Explorations: Bending Light**

Explore bending of light between two media with different indices of refraction. See how changing from air to water to glass changes the bending angle. Play with prisms of different shapes and make rainbows.

https://phet.colorado.edu/sims/html/bending-light/latest/bending-light_en.html

Section Summary

- The incident angle that produces an angle of refraction of 90° is called critical angle.
- Total internal reflection is a phenomenon that occurs at the boundary between two mediums, such that if the incident angle in the first medium is greater than the critical angle, then all the light is reflected back into that medium.
- Fiber optics involves the transmission of light down fibers of plastic or glass, applying the principle of total internal reflection.
- Endoscopes are used to explore the body through various orifices or minor incisions, based on the transmission of light through optical fibers.
- Cladding prevents light from being transmitted between fibers in a bundle.
- Diamonds sparkle due to total internal reflection coupled with a large index of refraction.

Conceptual Questions

Exercise:**Problem:**

A ring with a colorless gemstone is dropped into water. The gemstone becomes invisible when submerged. Can it be a diamond? Explain.

Exercise:**Problem:**

A high-quality diamond may be quite clear and colorless, transmitting all visible wavelengths with little absorption. Explain how it can sparkle with flashes of brilliant color when illuminated by white light.

Exercise:**Problem:**

Is it possible that total internal reflection plays a role in rainbows? Explain in terms of indices of refraction and angles, perhaps referring to [\[link\]](#). Some of us have seen the formation of a double rainbow. Is it physically possible to observe a triple rainbow?



Double rainbows are not a very common observance. (credit: InvictusOU812, Flickr)

Exercise:

Problem:

The most common type of mirage is an illusion that light from faraway objects is reflected by a pool of water that is not really there. Mirages are generally observed in deserts, when there is a hot layer of air near the ground. Given that the refractive index of air is lower for air at higher temperatures, explain how mirages can be formed.

Problems & Exercises**Exercise:****Problem:**

Verify that the critical angle for light going from water to air is 48.6° , as discussed at the end of [\[link\]](#), regarding the critical angle for light traveling in a polystyrene (a type of plastic) pipe surrounded by air.

Exercise:**Problem:**

(a) At the end of [\[link\]](#), it was stated that the critical angle for light going from diamond to air is 24.4° . Verify this. (b) What is the critical angle for light going from zircon to air?

Exercise:**Problem:**

An optical fiber uses flint glass clad with crown glass. What is the critical angle?

Solution:

66.3°

Exercise:

Problem:

At what minimum angle will you get total internal reflection of light traveling in water and reflected from ice?

Exercise:**Problem:**

Suppose you are using total internal reflection to make an efficient corner reflector. If there is air outside and the incident angle is 45.0° , what must be the minimum index of refraction of the material from which the reflector is made?

Solution:

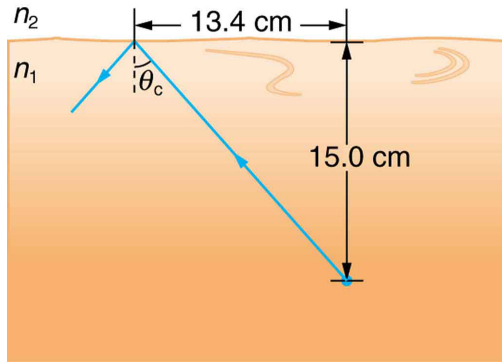
> 1.414

Exercise:**Problem:**

You can determine the index of refraction of a substance by determining its critical angle. (a) What is the index of refraction of a substance that has a critical angle of 68.4° when submerged in water? What is the substance, based on [\[link\]](#)? (b) What would the critical angle be for this substance in air?

Exercise:**Problem:**

A ray of light, emitted beneath the surface of an unknown liquid with air above it, undergoes total internal reflection as shown in [\[link\]](#). What is the index of refraction for the liquid and its likely identification?



A light ray inside a liquid strikes the surface at the critical angle and undergoes total internal reflection.

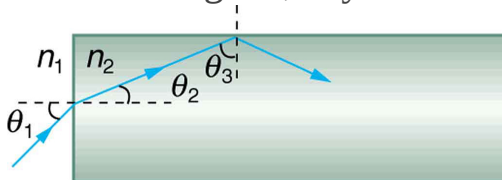
Solution:

1.50, benzene

Exercise:

Problem:

A light ray entering an optical fiber surrounded by air is first refracted and then reflected as shown in [\[link\]](#). Show that if the fiber is made from crown glass, any incident ray will be totally internally reflected.



A light ray enters the end of a fiber, the surface of which is perpendicular to its sides. Examine the conditions under which it

may be totally internally reflected.

Glossary

critical angle

incident angle that produces an angle of refraction of 90°

fiber optics

transmission of light down fibers of plastic or glass, applying the principle of total internal reflection

corner reflector

an object consisting of two mutually perpendicular reflecting surfaces, so that the light that enters is reflected back exactly parallel to the direction from which it came

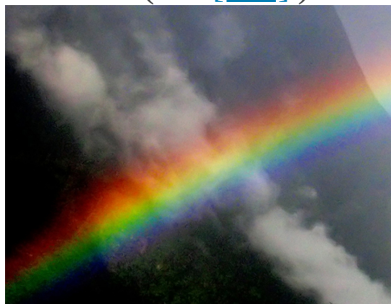
zircon

natural gemstone with a large index of refraction

Dispersion: The Rainbow and Prisms

- Explain the phenomenon of dispersion and discuss its advantages and disadvantages.

Everyone enjoys the spectacle of a rainbow glimmering against a dark stormy sky. How does sunlight falling on clear drops of rain get broken into the rainbow of colors we see? The same process causes white light to be broken into colors by a clear glass prism or a diamond. (See [\[link\]](#).)



(a)



(b)

The colors of the rainbow (a) and those produced by a prism (b) are identical.

(credit: Alfredo55, Wikimedia Commons; NASA)

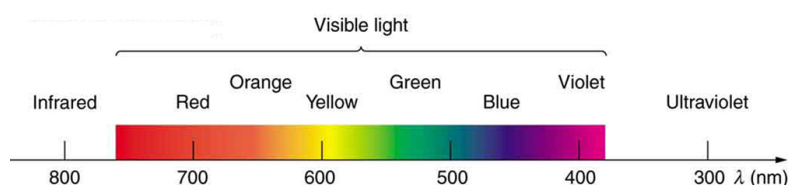
We see about six colors in a rainbow—red, orange, yellow, green, blue, and violet; sometimes indigo is listed, too. Those colors are associated with different wavelengths of light, as shown in [\[link\]](#). When our eye receives pure-wavelength light, we tend to see only one of the six colors, depending on wavelength. The thousands of other hues we can sense in other situations are our eye's response to various mixtures of wavelengths. White light, in particular, is a fairly uniform mixture of all visible wavelengths. Sunlight, considered to be white, actually appears to be a bit yellow because of its mixture of wavelengths, but it does contain all visible wavelengths. The sequence of colors in rainbows is the same sequence as the colors plotted versus wavelength in [\[link\]](#). What this implies is that white light is spread out according to

wavelength in a rainbow. **Dispersion** is defined as the spreading of white light into its full spectrum of wavelengths. More technically, dispersion occurs whenever there is a process that changes the direction of light in a manner that depends on wavelength. Dispersion, as a general phenomenon, can occur for any type of wave and always involves wavelength-dependent processes.

Note:

Dispersion

Dispersion is defined to be the spreading of white light into its full spectrum of wavelengths.



Even though rainbows are associated with seven colors, the rainbow is a continuous distribution of colors according to wavelengths.

Refraction is responsible for dispersion in rainbows and many other situations. The angle of refraction depends on the index of refraction, as we saw in [The Law of Refraction](#). We know that the index of refraction n depends on the medium. But for a given medium, n also depends on wavelength. (See [\[link\]](#). Note that, for a given medium, n increases as wavelength decreases and is greatest for violet light. Thus violet light is bent more than red light, as shown for a prism in [\[link\]\(b\)](#), and the light is dispersed into the same sequence of wavelengths as seen in [\[link\]](#) and [\[link\]](#).

Note:

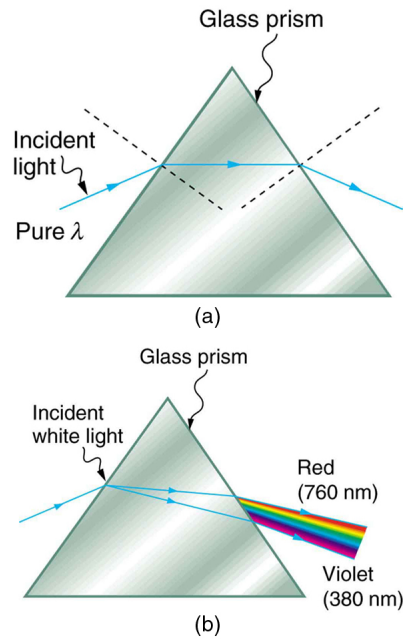
Making Connections: Dispersion

Any type of wave can exhibit dispersion. Sound waves, all types of electromagnetic waves, and water waves can be dispersed according to wavelength. Dispersion occurs whenever the speed of propagation depends on wavelength, thus separating and spreading out various wavelengths. Dispersion may require special circumstances and can result in spectacular displays such as in the production of a rainbow. This is also

true for sound, since all frequencies ordinarily travel at the same speed. If you listen to sound through a long tube, such as a vacuum cleaner hose, you can easily hear it is dispersed by interaction with the tube. Dispersion, in fact, can reveal a great deal about what the wave has encountered that disperses its wavelengths. The dispersion of electromagnetic radiation from outer space, for example, has revealed much about what exists between the stars—the so-called empty space.

Medium	Red (660 nm)	Orange (610 nm)	Yellow (580 nm)	Green (550 nm)	Blue (470 nm)	Violet (410 nm)
Water	1.331	1.332	1.333	1.335	1.338	1.342
Diamond	2.410	2.415	2.417	2.426	2.444	2.458
Glass, crown	1.512	1.514	1.518	1.519	1.524	1.530
Glass, flint	1.662	1.665	1.667	1.674	1.684	1.698
Polystyrene	1.488	1.490	1.492	1.493	1.499	1.506
Quartz, fused	1.455	1.456	1.458	1.459	1.462	1.468

Index of Refraction n in Selected Media at Various Wavelengths



(a) A pure wavelength of light falls onto a prism and is refracted at both surfaces. (b)

White light is dispersed by the prism (shown exaggerated).

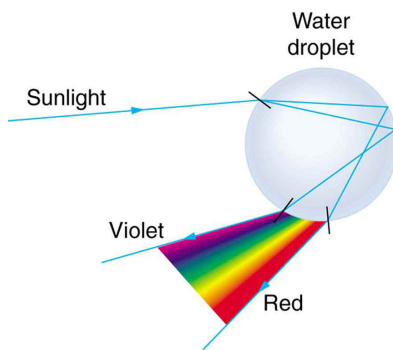
Since the index of refraction varies with wavelength, the angles of refraction vary with wavelength. A sequence of red to violet is produced, because the index of refraction increases steadily with decreasing wavelength.

Rainbows are produced by a combination of refraction and reflection. You may have noticed that you see a rainbow only when you look away from the sun. Light enters a drop of water and is reflected from the back of the drop, as shown in [\[link\]](#). The light is refracted both as it enters and as it leaves the drop. Since the index of refraction of water

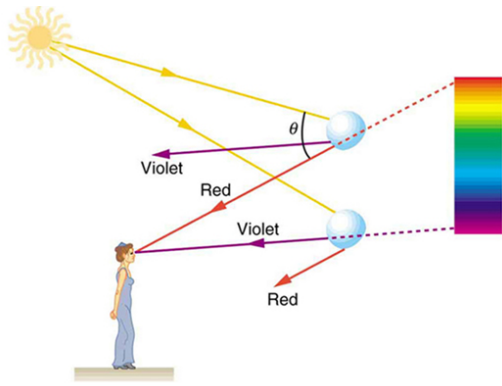
varies with wavelength, the light is dispersed, and a rainbow is observed, as shown in [\[link\]](#) (a). (There is no dispersion caused by reflection at the back surface, since the law of reflection does not depend on wavelength.) The actual rainbow of colors seen by an observer depends on the myriad of rays being refracted and reflected toward the observer's eyes from numerous drops of water. The effect is most spectacular when the background is dark, as in stormy weather, but can also be observed in waterfalls and lawn sprinklers. The arc of a rainbow comes from the need to be looking at a specific angle relative to the direction of the sun, as illustrated in [\[link\]](#) (b). (If there are two reflections of light within the water drop, another “secondary” rainbow is produced. This rare event produces an arc that lies above the primary rainbow arc—see [\[link\]](#) (c).)

Note:**Rainbows**

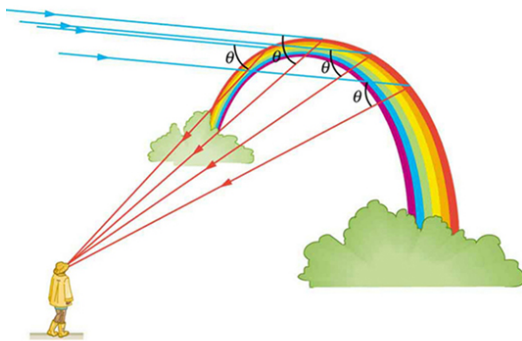
Rainbows are produced by a combination of refraction and reflection.



Part of the light falling on this water drop enters and is reflected from the back of the drop. This light is refracted and dispersed both as it enters and as it leaves the drop.



(a)



(b)



(c)

(a) Different colors emerge in different directions, and so you must look at different locations to see the various colors of a rainbow. (b) The arc of a rainbow results from the fact that a line between the observer and any point on the arc must make the correct angle with the parallel rays of sunlight to receive the refracted rays. (c)

Double rainbow. (credit:
Nicholas, Wikimedia
Commons)

Dispersion may produce beautiful rainbows, but it can cause problems in optical systems. White light used to transmit messages in a fiber is dispersed, spreading out in time and eventually overlapping with other messages. Since a laser produces a nearly pure wavelength, its light experiences little dispersion, an advantage over white light for transmission of information. In contrast, dispersion of electromagnetic waves coming to us from outer space can be used to determine the amount of matter they pass through. As with many phenomena, dispersion can be useful or a nuisance, depending on the situation and our human goals.

Note:

PhET Explorations: Geometric Optics

How does a lens form an image? See how light rays are refracted by a lens. Watch how the image changes when you adjust the focal length of the lens, move the object, move the lens, or move the screen.

https://phet.colorado.edu/sims/geometric-optics/geometric-optics_en.html

Section Summary

- The spreading of white light into its full spectrum of wavelengths is called dispersion.
- Rainbows are produced by a combination of refraction and reflection and involve the dispersion of sunlight into a continuous distribution of colors.
- Dispersion produces beautiful rainbows but also causes problems in certain optical systems.

Problems & Exercises

Exercise:

Problem:

(a) What is the ratio of the speed of red light to violet light in diamond, based on [link](#)? (b) What is this ratio in polystyrene? (c) Which is more dispersive?

Exercise:**Problem:**

A beam of white light goes from air into water at an incident angle of 75.0° . At what angles are the red (660 nm) and violet (410 nm) parts of the light refracted?

Solution:

46.5° , red; 46.0° , violet

Exercise:**Problem:**

By how much do the critical angles for red (660 nm) and violet (410 nm) light differ in a diamond surrounded by air?

Exercise:**Problem:**

(a) A narrow beam of light containing yellow (580 nm) and green (550 nm) wavelengths goes from polystyrene to air, striking the surface at a 30.0° incident angle. What is the angle between the colors when they emerge? (b) How far would they have to travel to be separated by 1.00 mm?

Solution:

(a) 0.043°

(b) 1.33 m

Exercise:**Problem:**

A parallel beam of light containing orange (610 nm) and violet (410 nm) wavelengths goes from fused quartz to water, striking the surface between them at a 60.0° incident angle. What is the angle between the two colors in water?

Exercise:**Problem:**

A ray of 610 nm light goes from air into fused quartz at an incident angle of 55.0° . At what incident angle must 470 nm light enter flint glass to have the same angle of refraction?

Solution:

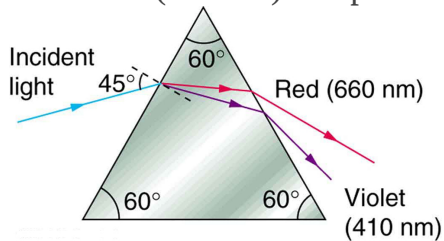
71.3°

Exercise:**Problem:**

A narrow beam of light containing red (660 nm) and blue (470 nm) wavelengths travels from air through a 1.00 cm thick flat piece of crown glass and back to air again. The beam strikes at a 30.0° incident angle. (a) At what angles do the two colors emerge? (b) By what distance are the red and blue separated when they emerge?

Exercise:**Problem:**

A narrow beam of white light enters a prism made of crown glass at a 45.0° incident angle, as shown in [\[link\]](#). At what angles, θ_R and θ_V , do the red (660 nm) and violet (410 nm) components of the light emerge from the prism?



This prism will disperse the white light into a rainbow of colors. The incident angle is 45.0°, and the angles at which the red and violet light emerge are θ_R and θ_V .

Solution:

53.5°, red; 55.2°, violet

Glossary

dispersion

spreading of white light into its full spectrum of wavelengths

rainbow

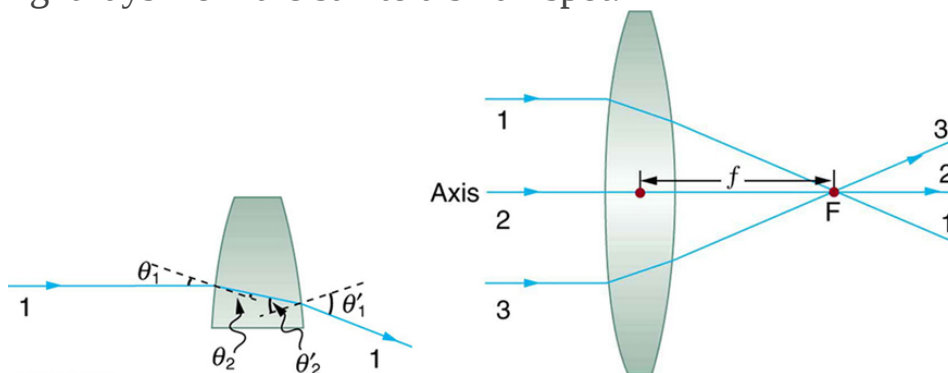
dispersion of sunlight into a continuous distribution of colors according to wavelength, produced by the refraction and reflection of sunlight by water droplets in the sky

Image Formation by Lenses

- List the rules for ray tracking for thin lenses.
- Illustrate the formation of images using the technique of ray tracking.
- Determine power of a lens given the focal length.

Lenses are found in a huge array of optical instruments, ranging from a simple magnifying glass to the eye to a camera's zoom lens. In this section, we will use the law of refraction to explore the properties of lenses and how they form images.

The word *lens* derives from the Latin word for a lentil bean, the shape of which is similar to the convex lens in [\[link\]](#). The convex lens shown has been shaped so that all light rays that enter it parallel to its axis cross one another at a single point on the opposite side of the lens. (The axis is defined to be a line normal to the lens at its center, as shown in [\[link\]](#).) Such a lens is called a **converging (or convex) lens** for the converging effect it has on light rays. An expanded view of the path of one ray through the lens is shown, to illustrate how the ray changes direction both as it enters and as it leaves the lens. Since the index of refraction of the lens is greater than that of air, the ray moves towards the perpendicular as it enters and away from the perpendicular as it leaves. (This is in accordance with the law of refraction.) Due to the lens's shape, light is thus bent toward the axis at both surfaces. The point at which the rays cross is defined to be the **focal point** F of the lens. The distance from the center of the lens to its focal point is defined to be the **focal length** f of the lens. [\[link\]](#) shows how a converging lens, such as that in a magnifying glass, can converge the nearly parallel light rays from the sun to a small spot.



Rays of light entering a converging lens parallel to its axis converge at its focal point F. (Ray 2 lies on the axis of the lens.) The distance from the center of the lens to the focal point is the lens's focal length f . An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Note:

Converging or Convex Lens

The lens in which light rays that enter it parallel to its axis cross one another at a single point on the opposite side with a converging effect is called converging lens.

Note:

Focal Point F

The point at which the light rays cross is called the focal point F of the lens.

Note:

Focal Length f

The distance from the center of the lens to its focal point is called focal length f .



Sunlight focused by a converging magnifying glass can burn paper. Light rays from the sun are nearly parallel and cross at the focal point of the lens. The more powerful the lens, the closer to the lens the rays will cross.

The greater effect a lens has on light rays, the more powerful it is said to be. For example, a powerful converging lens will focus parallel light rays closer to itself and will have a smaller focal length than a weak lens. The light will also focus into a smaller and more intense spot for a more powerful lens. The **power** P of a lens is defined to be the inverse of its focal length. In equation form, this is

Equation:

$$P = \frac{1}{f}.$$

Note:**Power P**

The **power** P of a lens is defined to be the inverse of its focal length. In equation form, this is

Equation:

$$P = \frac{1}{f}.$$

where f is the focal length of the lens, which must be given in meters (and not cm or mm). The power of a lens P has the unit diopters (D), provided that the focal length is given in meters. That is, $1 \text{ D} = 1/\text{m}$, or 1 m^{-1} .

(Note that this power (optical power, actually) is not the same as power in watts defined in [Work, Energy, and Energy Resources](#). It is a concept related to the effect of optical devices on light.) Optometrists prescribe common spectacles and contact lenses in units of diopters.

Example:**What is the Power of a Common Magnifying Glass?**

Suppose you take a magnifying glass out on a sunny day and you find that it concentrates sunlight to a small spot 8.00 cm away from the lens. What are the focal length and power of the lens?

Strategy

The situation here is the same as those shown in [\[link\]](#) and [\[link\]](#). The Sun is so far away that the Sun's rays are nearly parallel when they reach Earth. The magnifying glass is a convex (or converging) lens, focusing the nearly parallel rays of sunlight. Thus the focal length of the lens is the distance from the lens to the spot, and its power is the inverse of this distance (in m).

Solution

The focal length of the lens is the distance from the center of the lens to the spot, given to be 8.00 cm. Thus,

Equation:

$$f = 8.00 \text{ cm}.$$

To find the power of the lens, we must first convert the focal length to meters; then, we substitute this value into the equation for power. This gives

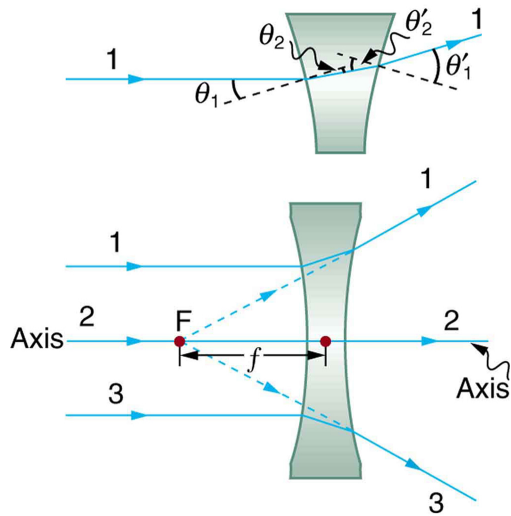
Equation:

$$P = \frac{1}{f} = \frac{1}{0.0800 \text{ m}} = 12.5 \text{ D.}$$

Discussion

This is a relatively powerful lens. The power of a lens in diopters should not be confused with the familiar concept of power in watts. It is an unfortunate fact that the word “power” is used for two completely different concepts. If you examine a prescription for eyeglasses, you will note lens powers given in diopters. If you examine the label on a motor, you will note energy consumption rate given as a power in watts.

[\[link\]](#) shows a concave lens and the effect it has on rays of light that enter it parallel to its axis (the path taken by ray 2 in the figure is the axis of the lens). The concave lens is a **diverging lens**, because it causes the light rays to bend away (diverge) from its axis. In this case, the lens has been shaped so that all light rays entering it parallel to its axis appear to originate from the same point, F , defined to be the focal point of a diverging lens. The distance from the center of the lens to the focal point is again called the focal length f of the lens. Note that the focal length and power of a diverging lens are defined to be negative. For example, if the distance to F in [\[link\]](#) is 5.00 cm, then the focal length is $f = -5.00 \text{ cm}$ and the power of the lens is $P = -20 \text{ D}$. An expanded view of the path of one ray through the lens is shown in the figure to illustrate how the shape of the lens, together with the law of refraction, causes the ray to follow its particular path and be diverged.

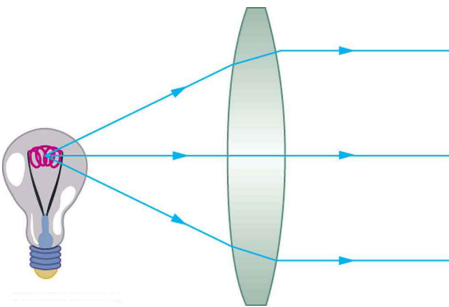


Rays of light entering a diverging lens parallel to its axis are diverged, and all appear to originate at its focal point F. The dashed lines are not rays—they indicate the directions from which the rays appear to come. The focal length f of a diverging lens is negative. An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Note:
Diverging Lens

A lens that causes the light rays to bend away from its axis is called a diverging lens.

As noted in the initial discussion of the law of refraction in [The Law of Refraction](#), the paths of light rays are exactly reversible. This means that the direction of the arrows could be reversed for all of the rays in [\[link\]](#) and [\[link\]](#). For example, if a point light source is placed at the focal point of a convex lens, as shown in [\[link\]](#), parallel light rays emerge from the other side.



A small light source, like a light bulb filament, placed at the focal point of a convex lens, results in parallel rays of light emerging from the other side. The paths are exactly the reverse of those shown in [\[link\]](#). This technique is used in lighthouses and sometimes in traffic lights to produce a directional beam of light from a source that emits light in all directions.

Ray Tracing and Thin Lenses

Ray tracing is the technique of determining or following (tracing) the paths that light rays take. For rays passing through matter, the law of refraction is used to trace the paths. Here we use ray tracing to help us understand the action of lenses in situations ranging from forming images on film to magnifying small print to correcting nearsightedness. While ray tracing for complicated lenses, such as those found in sophisticated cameras, may require computer techniques, there is a set of simple rules for tracing rays through thin lenses. A **thin lens** is defined to be one whose thickness allows rays to refract, as illustrated in [\[link\]](#), but does not allow properties such as dispersion and aberrations. An ideal thin lens has two refracting surfaces but the lens is thin enough to assume that light rays bend only once. A thin symmetrical lens has two focal points, one on either side and both at the same distance from the lens. (See [\[link\]](#).) Another important characteristic of a thin lens is that light rays through its center are deflected by a negligible amount, as seen in [\[link\]](#).

Note:

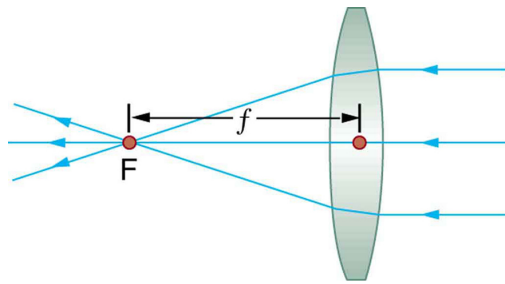
Thin Lens

A thin lens is defined to be one whose thickness allows rays to refract but does not allow properties such as dispersion and aberrations.

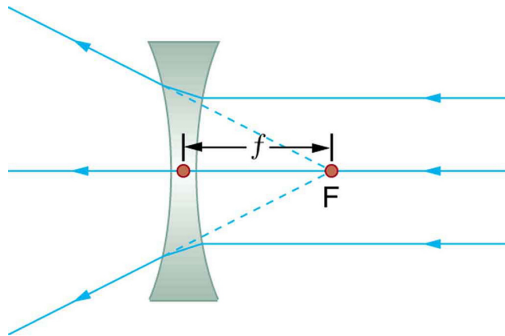
Note:

Take-Home Experiment: A Visit to the Optician

Look through your eyeglasses (or those of a friend) backward and forward and comment on whether they act like thin lenses.

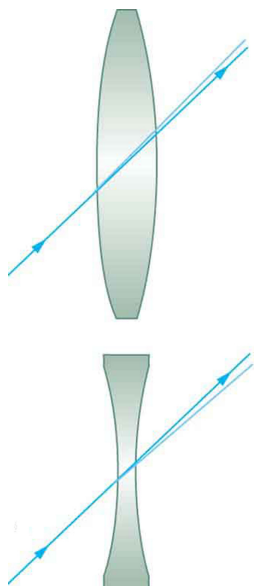


(a)



(b)

Thin lenses have the same focal length on either side. (a) Parallel light rays entering a converging lens from the right cross at its focal point on the left. (b) Parallel light rays entering a diverging lens from the right seem to come from the focal point on the right.



The light ray through the center of a thin lens is deflected by a negligible amount and is assumed to emerge parallel to its original path (shown as a shaded line).

Using paper, pencil, and a straight edge, ray tracing can accurately describe the operation of a lens. The rules for ray tracing for thin lenses are based on the illustrations already discussed:

1. A ray entering a converging lens parallel to its axis passes through the focal point F of the lens on the other side. (See rays 1 and 3 in [\[link\]](#).)
2. A ray entering a diverging lens parallel to its axis seems to come from the focal point F . (See rays 1 and 3 in [\[link\]](#).)
3. A ray passing through the center of either a converging or a diverging lens does not change direction. (See [\[link\]](#), and see ray 2 in [\[link\]](#) and [\[link\]](#).)
4. A ray entering a converging lens through its focal point exits parallel to its axis. (The reverse of rays 1 and 3 in [\[link\]](#).)
5. A ray that enters a diverging lens by heading toward the focal point on the opposite side exits parallel to the axis. (The reverse of rays 1 and 3 in [\[link\]](#).)

Note:

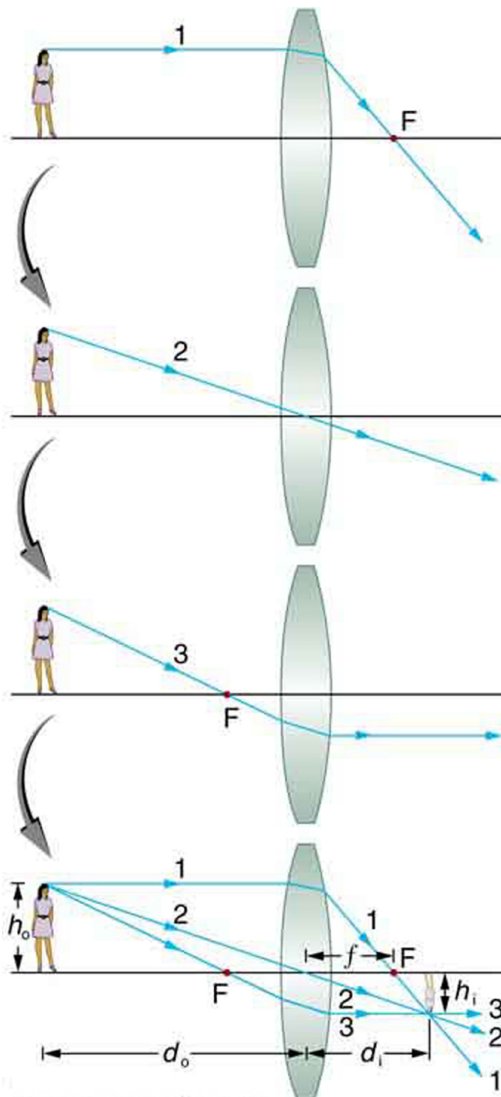
Rules for Ray Tracing

1. A ray entering a converging lens parallel to its axis passes through the focal point F of the lens on the other side.
2. A ray entering a diverging lens parallel to its axis seems to come from the focal point F .
3. A ray passing through the center of either a converging or a diverging lens does not change direction.
4. A ray entering a converging lens through its focal point exits parallel to its axis.
5. A ray that enters a diverging lens by heading toward the focal point on the opposite side exits parallel to the axis.

Image Formation by Thin Lenses

In some circumstances, a lens forms an obvious image, such as when a movie projector casts an image onto a screen. In other cases, the image is less obvious. Where, for example, is the image formed by eyeglasses? We will use ray tracing for thin lenses to illustrate how they form images, and we will develop equations to describe the image formation quantitatively.

Consider an object some distance away from a converging lens, as shown in [\[link\]](#). To find the location and size of the image formed, we trace the paths of selected light rays originating from one point on the object, in this case the top of the person's head. The figure shows three rays from the top of the object that can be traced using the ray tracing rules given above. (Rays leave this point going in many directions, but we concentrate on only a few with paths that are easy to trace.) The first ray is one that enters the lens parallel to its axis and passes through the focal point on the other side (rule 1). The second ray passes through the center of the lens without changing direction (rule 3). The third ray passes through the nearer focal point on its way into the lens and leaves the lens parallel to its axis (rule 4). The three rays cross at the same point on the other side of the lens. The image of the top of the person's head is located at this point. All rays that come from the same point on the top of the person's head are refracted in such a way as to cross at the point shown. Rays from another point on the object, such as her belt buckle, will also cross at another common point, forming a complete image, as shown. Although three rays are traced in [\[link\]](#), only two are necessary to locate the image. It is best to trace rays for which there are simple ray tracing rules. Before applying ray tracing to other situations, let us consider the example shown in [\[link\]](#) in more detail.



Ray tracing is used to locate the image formed by a lens. Rays originating from the same point on the object are traced—the three chosen rays each follow one of the rules for ray tracing, so that their paths are easy to determine. The image is located at the point where the rays

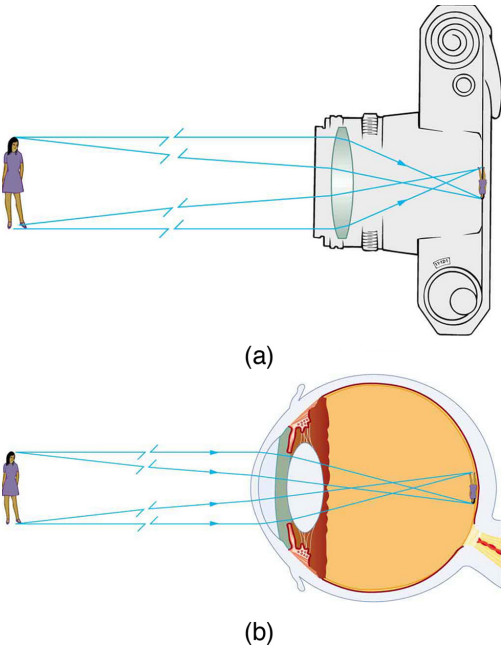
cross. In this case, a real image—one that can be projected on a screen—is formed.

The image formed in [\[link\]](#) is a **real image**, meaning that it can be projected. That is, light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye, for example. [\[link\]](#) shows how such an image would be projected onto film by a camera lens. This figure also shows how a real image is projected onto the retina by the lens of an eye. Note that the image is there whether it is projected onto a screen or not.

Note:

Real Image

The image in which light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye is called a real image.



Real images can be projected. (a) A real image of the person is projected onto film. (b) The converging nature of the multiple surfaces that make up the eye result in the projection of a real image on the retina.

Several important distances appear in [\[link\]](#). We define d_o to be the object distance, the distance of an object from the center of a lens. **Image distance** d_i is defined to be the distance of the image from the center of a lens. The height of the object and height of the image are given the symbols h_o and h_i , respectively. Images that appear upright relative to the object have heights that are positive and those that are inverted have negative heights. Using the rules of ray tracing and making a scale drawing with paper and pencil, like that in [\[link\]](#), we can accurately describe the location and size of an image. But the real benefit of ray tracing is in visualizing how images are formed in a variety of situations. To obtain numerical information, we use a pair of

equations that can be derived from a geometric analysis of ray tracing for thin lenses. The **thin lens equations** are

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

and

Equation:

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m.$$

We define the ratio of image height to object height (h_i/h_o) to be the **magnification** m . (The minus sign in the equation above will be discussed shortly.) The thin lens equations are broadly applicable to all situations involving thin lenses (and “thin” mirrors, as we will see later). We will explore many features of image formation in the following worked examples.

Note:

Image Distance

The distance of the image from the center of the lens is called image distance.

Note:

Thin Lens Equations and Magnification

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

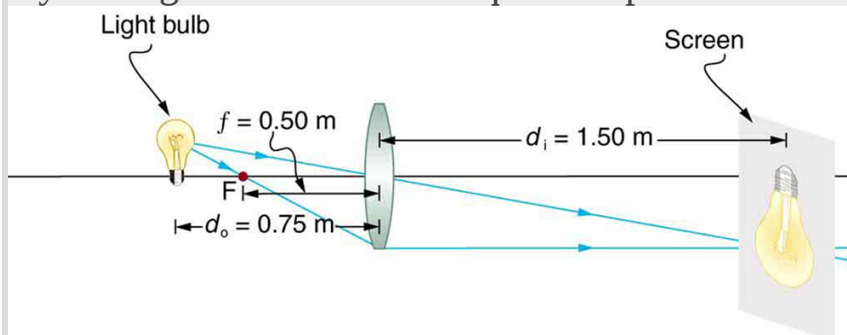
Equation:

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m$$

Example:

Finding the Image of a Light Bulb Filament by Ray Tracing and by the Thin Lens Equations

A clear glass light bulb is placed 0.750 m from a convex lens having a 0.500 m focal length, as shown in [\[link\]](#). Use ray tracing to get an approximate location for the image. Then use the thin lens equations to calculate (a) the location of the image and (b) its magnification. Verify that ray tracing and the thin lens equations produce consistent results.



A light bulb placed 0.750 m from a lens having a 0.500 m focal length produces a real image on a poster board as discussed in the example above. Ray tracing predicts the image location and size.

Strategy and Concept

Since the object is placed farther away from a converging lens than the focal length of the lens, this situation is analogous to those illustrated in [\[link\]](#) and [\[link\]](#). Ray tracing to scale should produce similar results for d_i . Numerical solutions for d_i and m can be obtained using the thin lens equations, noting that $d_o = 0.750 \text{ m}$ and $f = 0.500 \text{ m}$.

Solutions (Ray tracing)

The ray tracing to scale in [\[link\]](#) shows two rays from a point on the bulb's filament crossing about 1.50 m on the far side of the lens. Thus the image distance d_i is about 1.50 m. Similarly, the image height based on ray tracing is greater than the object height by about a factor of 2, and the image is inverted. Thus m is about -2 . The minus sign indicates that the image is inverted.

The thin lens equations can be used to find d_i from the given information:

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}.$$

Rearranging to isolate d_i gives

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}.$$

Entering known quantities gives a value for $1/d_i$:

Equation:

$$\frac{1}{d_i} = \frac{1}{0.500 \text{ m}} - \frac{1}{0.750 \text{ m}} = \frac{0.667}{\text{m}}.$$

This must be inverted to find d_i :

Equation:

$$d_i = \frac{\text{m}}{0.667} = 1.50 \text{ m}.$$

Note that another way to find d_i is to rearrange the equation:

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}.$$

This yields the equation for the image distance as:

Equation:

$$d_i = \frac{fd_o}{d_o - f}.$$

Note that there is no inverting here.

The thin lens equations can be used to find the magnification m , since both d_i and d_o are known. Entering their values gives

Equation:

$$m = -\frac{d_i}{d_o} = -\frac{1.50 \text{ m}}{0.750 \text{ m}} = -2.00.$$

Discussion

Note that the minus sign causes the magnification to be negative when the image is inverted. Ray tracing and the use of the thin lens equations produce consistent results. The thin lens equations give the most precise results, being limited only by the accuracy of the given information. Ray tracing is limited by the accuracy with which you can draw, but it is highly useful both conceptually and visually.

Real images, such as the one considered in the previous example, are formed by converging lenses whenever an object is farther from the lens than its focal length. This is true for movie projectors, cameras, and the eye. We shall refer to these as *case 1* images. A case 1 image is formed when $d_o > f$ and f is positive, as in [\[link\]\(a\)](#). (A summary of the three cases or types of image formation appears at the end of this section.)

A different type of image is formed when an object, such as a person's face, is held close to a convex lens. The image is upright and larger than the object, as seen in [\[link\]\(b\)](#), and so the lens is called a magnifier. If you slowly pull the magnifier away from the face, you will see that the magnification steadily increases until the image begins to blur. Pulling the magnifier even farther away produces an inverted image as seen in [\[link\]\(a\)](#). The distance at which the image blurs, and beyond which it inverts, is the focal length of the lens. To use a convex lens as a magnifier, the object

must be closer to the converging lens than its focal length. This is called a *case 2* image. A case 2 image is formed when $d_o < f$ and f is positive.



(a)



(b)

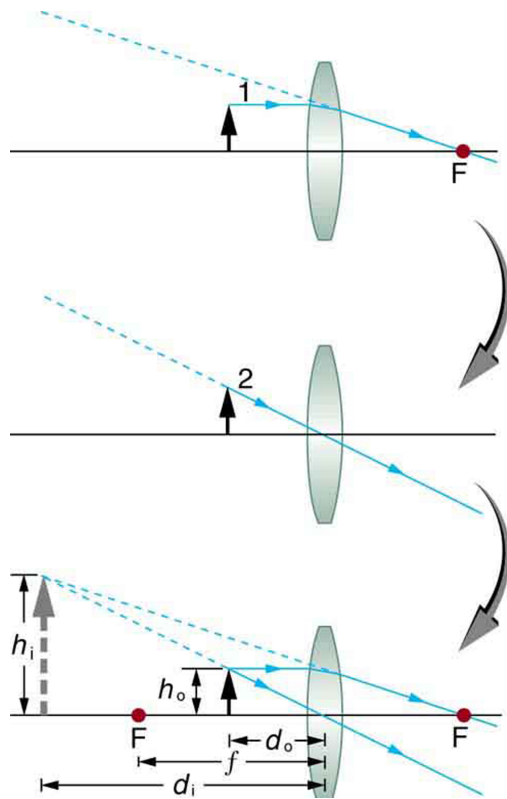
(a) When a converging lens is held farther away from the face than the lens's focal length, an inverted image is formed. This is a case 1 image. Note that the image is in focus but the face is not, because the image is much closer to the camera taking this photograph than the face. (credit:

DaMongMan, Flickr)

(b) A magnified image

of a face is produced
by placing it closer to
the converging lens
than its focal length.
This is a case 2 image.
(credit: Casey Fleser,
Flickr)

[\[link\]](#) uses ray tracing to show how an image is formed when an object is held closer to a converging lens than its focal length. Rays coming from a common point on the object continue to diverge after passing through the lens, but all appear to originate from a point at the location of the image. The image is on the same side of the lens as the object and is farther away from the lens than the object. This image, like all case 2 images, cannot be projected and, hence, is called a **virtual image**. Light rays only appear to originate at a virtual image; they do not actually pass through that location in space. A screen placed at the location of a virtual image will receive only diffuse light from the object, not focused rays from the lens. Additionally, a screen placed on the opposite side of the lens will receive rays that are still diverging, and so no image will be projected on it. We can see the magnified image with our eyes, because the lens of the eye converges the rays into a real image projected on our retina. Finally, we note that a virtual image is upright and larger than the object, meaning that the magnification is positive and greater than 1.



Ray tracing predicts the image location and size for an object held closer to a converging lens than its focal length. Ray 1 enters parallel to the axis and exits through the focal point on the opposite side, while ray 2 passes through the center of the lens without changing path. The two rays continue to diverge on the other side of the lens, but both appear to come from a common point, locating the upright, magnified,

virtual image. This is a case 2 image.

Note:

Virtual Image

An image that is on the same side of the lens as the object and cannot be projected on a screen is called a virtual image.

Example:

Image Produced by a Magnifying Glass

Suppose the book page in [\[link\]](#) (a) is held 7.50 cm from a convex lens of focal length 10.0 cm, such as a typical magnifying glass might have. What magnification is produced?

Strategy and Concept

We are given that $d_o = 7.50$ cm and $f = 10.0$ cm, so we have a situation where the object is placed closer to the lens than its focal length. We therefore expect to get a case 2 virtual image with a positive magnification that is greater than 1. Ray tracing produces an image like that shown in [\[link\]](#), but we will use the thin lens equations to get numerical solutions in this example.

Solution

To find the magnification m , we try to use magnification equation, $m = -d_i/d_o$. We do not have a value for d_i , so that we must first find the location of the image using lens equation. (The procedure is the same as followed in the preceding example, where d_o and f were known.)

Rearranging the magnification equation to isolate d_i gives

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}.$$

Entering known values, we obtain a value for $1/d_i$:

Equation:

$$\frac{1}{d_i} = \frac{1}{10.0 \text{ cm}} - \frac{1}{7.50 \text{ cm}} = \frac{-0.0333}{\text{cm}}.$$

This must be inverted to find d_i :

Equation:

$$d_i = -\frac{\text{cm}}{0.0333} = -30.0 \text{ cm}.$$

Now the thin lens equation can be used to find the magnification m , since both d_i and d_o are known. Entering their values gives

Equation:

$$m = -\frac{d_i}{d_o} = -\frac{-30.0 \text{ cm}}{7.50 \text{ cm}} = 4.00.$$

Discussion

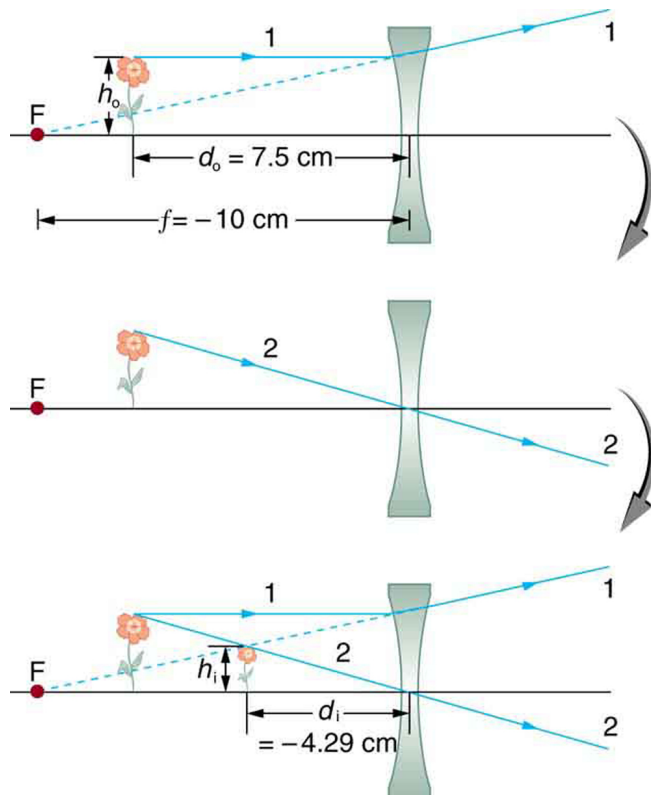
A number of results in this example are true of all case 2 images, as well as being consistent with [\[link\]](#). Magnification is indeed positive (as predicted), meaning the image is upright. The magnification is also greater than 1, meaning that the image is larger than the object—in this case, by a factor of 4. Note that the image distance is negative. This means the image is on the same side of the lens as the object. Thus the image cannot be projected and is virtual. (Negative values of d_i occur for virtual images.) The image is farther from the lens than the object, since the image distance is greater in magnitude than the object distance. The location of the image is not obvious when you look through a magnifier. In fact, since the image is bigger than the object, you may think the image is closer than the object. But the image is farther away, a fact that is useful in correcting farsightedness, as we shall see in a later section.

A third type of image is formed by a diverging or concave lens. Try looking through eyeglasses meant to correct nearsightedness. (See [\[link\]](#).) You will see an image that is upright but smaller than the object. This means that the magnification is positive but less than 1. The ray diagram in [\[link\]](#) shows that the image is on the same side of the lens as the object and, hence,

cannot be projected—it is a virtual image. Note that the image is closer to the lens than the object. This is a *case 3* image, formed for any object by a negative focal length or diverging lens.



A car viewed through a concave or diverging lens looks upright. This is a case 3 image. (credit: Daniel Oines, Flickr)



Ray tracing predicts the image location and size for a concave or diverging lens. Ray 1 enters parallel to the axis and is bent so that it appears to originate from the focal point. Ray 2 passes through the center of the lens without changing path. The two rays appear to come from a common point, locating the upright image. This is a case 3 image, which is closer to the lens than the object and smaller in height.

Example:

Image Produced by a Concave Lens

Suppose an object such as a book page is held 7.50 cm from a concave lens of focal length -10.0 cm. Such a lens could be used in eyeglasses to correct pronounced nearsightedness. What magnification is produced?

Strategy and Concept

This example is identical to the preceding one, except that the focal length is negative for a concave or diverging lens. The method of solution is thus the same, but the results are different in important ways.

Solution

To find the magnification m , we must first find the image distance d_i using thin lens equation

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o},$$

or its alternative rearrangement

Equation:

$$d_i = \frac{fd_o}{d_o - f}.$$

We are given that $f = -10.0$ cm and $d_o = 7.50$ cm. Entering these yields a value for $1/d_i$:

Equation:

$$\frac{1}{d_i} = \frac{1}{-10.0 \text{ cm}} - \frac{1}{7.50 \text{ cm}} = \frac{-0.2333}{\text{cm}}.$$

This must be inverted to find d_i :

Equation:

$$d_i = -\frac{\text{cm}}{0.2333} = -4.29 \text{ cm}.$$

Or

Equation:

$$d_i = \frac{(7.5)(-10)}{(7.5 - (-10))} = -75/17.5 = -4.29 \text{ cm.}$$

Now the magnification equation can be used to find the magnification m , since both d_i and d_o are known. Entering their values gives

Equation:

$$m = -\frac{d_i}{d_o} = -\frac{-4.29 \text{ cm}}{7.50 \text{ cm}} = 0.571.$$

Discussion

A number of results in this example are true of all case 3 images, as well as being consistent with [\[link\]](#). Magnification is positive (as predicted), meaning the image is upright. The magnification is also less than 1, meaning the image is smaller than the object—in this case, a little over half its size. The image distance is negative, meaning the image is on the same side of the lens as the object. (The image is virtual.) The image is closer to the lens than the object, since the image distance is smaller in magnitude than the object distance. The location of the image is not obvious when you look through a concave lens. In fact, since the image is smaller than the object, you may think it is farther away. But the image is closer than the object, a fact that is useful in correcting nearsightedness, as we shall see in a later section.

[\[link\]](#) summarizes the three types of images formed by single thin lenses. These are referred to as case 1, 2, and 3 images. Convex (converging) lenses can form either real or virtual images (cases 1 and 2, respectively), whereas concave (diverging) lenses can form only virtual images (always case 3). Real images are always inverted, but they can be either larger or smaller than the object. For example, a slide projector forms an image larger than the slide, whereas a camera makes an image smaller than the object being photographed. Virtual images are always upright and cannot be projected. Virtual images are larger than the object only in case 2, where a convex lens is used. The virtual image produced by a concave lens is

always smaller than the object—a case 3 image. We can see and photograph virtual images only by using an additional lens to form a real image.

Type	Formed when	Image type	d_i	m
Case 1	f positive, $d_o > f$	real	positive	negative
Case 2	f positive, $d_o < f$	virtual	negative	positive $m > 1$
Case 3	f negative	virtual	negative	positive $m < 1$

Three Types of Images Formed By Thin Lenses

In [Image Formation by Mirrors](#), we shall see that mirrors can form exactly the same types of images as lenses.

Note:

Take-Home Experiment: Concentrating Sunlight

Find several lenses and determine whether they are converging or diverging. In general those that are thicker near the edges are diverging and those that are thicker near the center are converging. On a bright sunny day take the converging lenses outside and try focusing the sunlight onto a piece of paper. Determine the focal lengths of the lenses. Be careful because the paper may start to burn, depending on the type of lens you have selected.

Problem-Solving Strategies for Lenses

Step 1. Examine the situation to determine that image formation by a lens is involved.

Step 2. Determine whether ray tracing, the thin lens equations, or both are to be employed. A sketch is very useful even if ray tracing is not specifically required by the problem. Write symbols and values on the sketch.

Step 3. Identify exactly what needs to be determined in the problem (identify the unknowns).

Step 4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is helpful to determine whether the situation involves a case 1, 2, or 3 image. While these are just names for types of images, they have certain characteristics (given in [\[link\]](#)) that can be of great use in solving problems.

Step 5. If ray tracing is required, use the ray tracing rules listed near the beginning of this section.

Step 6. Most quantitative problems require the use of the thin lens equations. These are solved in the usual manner by substituting knowns and solving for unknowns. Several worked examples serve as guides.

Step 7. Check to see if the answer is reasonable: Does it make sense? If you have identified the type of image (case 1, 2, or 3), you should assess whether your answer is consistent with the type of image, magnification, and so on.

Note:

Misconception Alert

We do not realize that light rays are coming from every part of the object, passing through every part of the lens, and all can be used to form the final image.

We generally feel the entire lens, or mirror, is needed to form an image. Actually, half a lens will form the same, though a fainter, image.

Section Summary

- Light rays entering a converging lens parallel to its axis cross one another at a single point on the opposite side.
- For a converging lens, the focal point is the point at which converging light rays cross; for a diverging lens, the focal point is the point from which diverging light rays appear to originate.
- The distance from the center of the lens to its focal point is called the focal length f .
- Power P of a lens is defined to be the inverse of its focal length, $P = \frac{1}{f}$.
- A lens that causes the light rays to bend away from its axis is called a diverging lens.
- Ray tracing is the technique of graphically determining the paths that light rays take.
- The image in which light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye is called a real image.
- Thin lens equations are $\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$ and $\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m$ (magnification).

- The distance of the image from the center of the lens is called image distance.
- An image that is on the same side of the lens as the object and cannot be projected on a screen is called a virtual image.

Conceptual Questions

Exercise:

Problem:

It can be argued that a flat piece of glass, such as in a window, is like a lens with an infinite focal length. If so, where does it form an image? That is, how are d_i and d_o related?

Exercise:

Problem:

You can often see a reflection when looking at a sheet of glass, particularly if it is darker on the other side. Explain why you can often see a double image in such circumstances.

Exercise:

Problem:

When you focus a camera, you adjust the distance of the lens from the film. If the camera lens acts like a thin lens, why can it not be a fixed distance from the film for both near and distant objects?

Exercise:

Problem:

A thin lens has two focal points, one on either side, at equal distances from its center, and should behave the same for light entering from either side. Look through your eyeglasses (or those of a friend) backward and forward and comment on whether they are thin lenses.

Exercise:

Problem:

Will the focal length of a lens change when it is submerged in water? Explain.

Problems & Exercises**Exercise:****Problem:**

What is the power in diopters of a camera lens that has a 50.0 mm focal length?

Exercise:**Problem:**

Your camera's zoom lens has an adjustable focal length ranging from 80.0 to 200 mm. What is its range of powers?

Solution:

5.00 to 12.5 D

Exercise:**Problem:**

What is the focal length of 1.75 D reading glasses found on the rack in a pharmacy?

Exercise:**Problem:**

You note that your prescription for new eyeglasses is -4.50 D. What will their focal length be?

Solution:

-0.222 m

Exercise:

Problem:

How far from the lens must the film in a camera be, if the lens has a 35.0 mm focal length and is being used to photograph a flower 75.0 cm away? Explicitly show how you follow the steps in the Problem-Solving Strategy for lenses.

Exercise:

Problem:

A certain slide projector has a 100 mm focal length lens. (a) How far away is the screen, if a slide is placed 103 mm from the lens and produces a sharp image? (b) If the slide is 24.0 by 36.0 mm, what are the dimensions of the image? Explicitly show how you follow the steps in the Problem-Solving Strategy for lenses.

Solution:

(a) 3.43 m

(b) 0.800 by 1.20 m

Exercise:

Problem:

A doctor examines a mole with a 15.0 cm focal length magnifying glass held 13.5 cm from the mole (a) Where is the image? (b) What is its magnification? (c) How big is the image of a 5.00 mm diameter mole?

Solution:

(a) -1.35 m (on the object side of the lens).

(b) $+10.0$

(c) 5.00 cm

Exercise:

Problem:

How far from a piece of paper must you hold your father's 2.25 D reading glasses to try to burn a hole in the paper with sunlight?

Solution:

44.4 cm

Exercise:

Problem:

A camera with a 50.0 mm focal length lens is being used to photograph a person standing 3.00 m away. (a) How far from the lens must the film be? (b) If the film is 36.0 mm high, what fraction of a 1.75 m tall person will fit on it? (c) Discuss how reasonable this seems, based on your experience in taking or posing for photographs.

Exercise:

Problem:

A camera lens used for taking close-up photographs has a focal length of 22.0 mm. The farthest it can be placed from the film is 33.0 mm. (a) What is the closest object that can be photographed? (b) What is the magnification of this closest object?

Solution:

(a) 6.60 cm

(b) -0.333

Exercise:

Problem:

Suppose your 50.0 mm focal length camera lens is 51.0 mm away from the film in the camera. (a) How far away is an object that is in focus? (b) What is the height of the object if its image is 2.00 cm high?

Exercise:**Problem:**

(a) What is the focal length of a magnifying glass that produces a magnification of 3.00 when held 5.00 cm from an object, such as a rare coin? (b) Calculate the power of the magnifier in diopters. (c) Discuss how this power compares to those for store-bought reading glasses (typically 1.0 to 4.0 D). Is the magnifier's power greater, and should it be?

Solution:

(a) +7.50 cm

(b) 13.3 D

(c) Much greater

Exercise:**Problem:**

What magnification will be produced by a lens of power -4.00 D (such as might be used to correct myopia) if an object is held 25.0 cm away?

Exercise:**Problem:**

In [\[link\]](#), the magnification of a book held 7.50 cm from a 10.0 cm focal length lens was found to be 3.00. (a) Find the magnification for the book when it is held 8.50 cm from the magnifier. (b) Do the same for when it is held 9.50 cm from the magnifier. (c) Comment on the trend in m as the object distance increases as in these two calculations.

Solution:

(a) +6.67

(b) +20.0

(c) The magnification increases without limit (to infinity) as the object distance increases to the limit of the focal distance.

Exercise:**Problem:**

Suppose a 200 mm focal length telephoto lens is being used to photograph mountains 10.0 km away. (a) Where is the image? (b) What is the height of the image of a 1000 m high cliff on one of the mountains?

Exercise:**Problem:**

A camera with a 100 mm focal length lens is used to photograph the sun and moon. What is the height of the image of the sun on the film, given the sun is 1.40×10^6 km in diameter and is 1.50×10^8 km away?

Solution:

−0.933 mm

Exercise:**Problem:**

Combine thin lens equations to show that the magnification for a thin lens is determined by its focal length and the object distance and is given by $m = f/(f - d_o)$.

Glossary

converging lens

a convex lens in which light rays that enter it parallel to its axis converge at a single point on the opposite side

diverging lens

a concave lens in which light rays that enter it parallel to its axis bend away (diverge) from its axis

focal point

for a converging lens or mirror, the point at which converging light rays cross; for a diverging lens or mirror, the point from which diverging light rays appear to originate

focal length

distance from the center of a lens or curved mirror to its focal point

magnification

ratio of image height to object height

power

inverse of focal length

real image

image that can be projected

virtual image

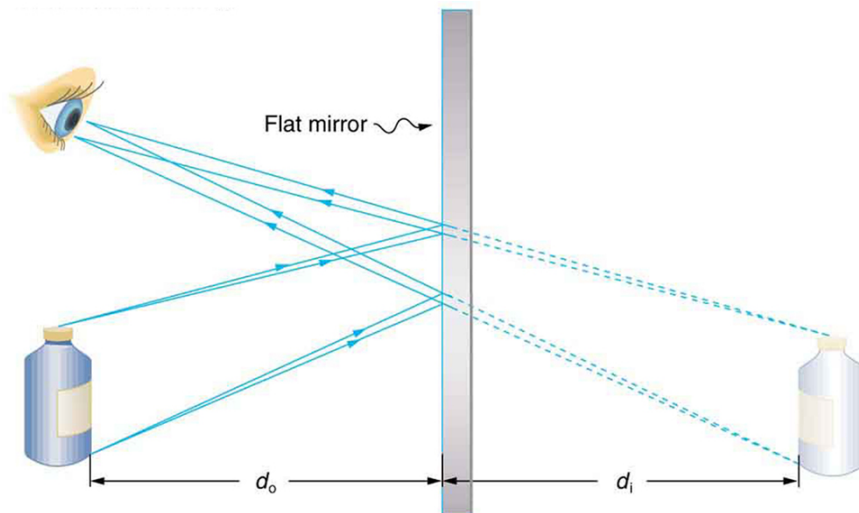
image that cannot be projected

Image Formation by Mirrors

- Illustrate image formation in a flat mirror.
- Explain with ray diagrams the formation of an image using spherical mirrors.
- Determine focal length and magnification given radius of curvature, distance of object and image.

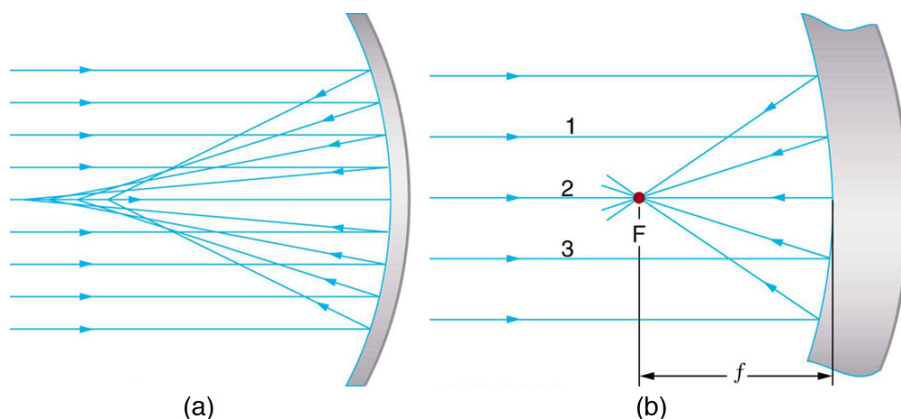
We only have to look as far as the nearest bathroom to find an example of an image formed by a mirror. Images in flat mirrors are the same size as the object and are located behind the mirror. Like lenses, mirrors can form a variety of images. For example, dental mirrors may produce a magnified image, just as makeup mirrors do. Security mirrors in shops, on the other hand, form images that are smaller than the object. We will use the law of reflection to understand how mirrors form images, and we will find that mirror images are analogous to those formed by lenses.

[\[link\]](#) helps illustrate how a flat mirror forms an image. Two rays are shown emerging from the same point, striking the mirror, and being reflected into the observer's eye. The rays can diverge slightly, and both still get into the eye. If the rays are extrapolated backward, they seem to originate from a common point behind the mirror, locating the image. (The paths of the reflected rays into the eye are the same as if they had come directly from that point behind the mirror.) Using the law of reflection—the angle of reflection equals the angle of incidence—we can see that the image and object are the same distance from the mirror. This is a virtual image, since it cannot be projected—the rays only appear to originate from a common point behind the mirror. Obviously, if you walk behind the mirror, you cannot see the image, since the rays do not go there. But in front of the mirror, the rays behave exactly as if they had come from behind the mirror, so that is where the image is situated.



Two sets of rays from common points on an object are reflected by a flat mirror into the eye of an observer. The reflected rays seem to originate from behind the mirror, locating the virtual image.

Now let us consider the focal length of a mirror—for example, the concave spherical mirrors in [\[link\]](#). Rays of light that strike the surface follow the law of reflection. For a mirror that is large compared with its radius of curvature, as in [\[link\]](#)(a), we see that the reflected rays do not cross at the same point, and the mirror does not have a well-defined focal point. If the mirror had the shape of a parabola, the rays would all cross at a single point, and the mirror would have a well-defined focal point. But parabolic mirrors are much more expensive to make than spherical mirrors. The solution is to use a mirror that is small compared with its radius of curvature, as shown in [\[link\]](#)(b). (This is the mirror equivalent of the thin lens approximation.) To a very good approximation, this mirror has a well-defined focal point at F that is the focal distance f from the center of the mirror. The focal length f of a concave mirror is positive, since it is a converging mirror.



(a) Parallel rays reflected from a large spherical mirror do not all cross at a common point. (b) If a spherical mirror is small compared with its radius of curvature, parallel rays are focused to a common point. The distance of the focal point from the center of the mirror is its focal length f . Since this mirror is converging, it has a positive focal length.

Just as for lenses, the shorter the focal length, the more powerful the mirror; thus, $P = 1/f$ for a mirror, too. A more strongly curved mirror has a shorter focal length and a greater power. Using the law of reflection and some simple trigonometry, it can be shown that the focal length is half the radius of curvature, or

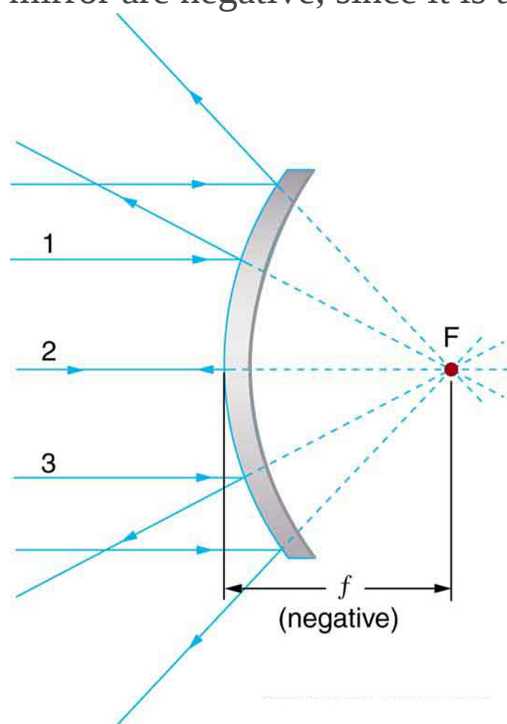
Equation:

$$f = \frac{R}{2},$$

where R is the radius of curvature of a spherical mirror. The smaller the radius of curvature, the smaller the focal length and, thus, the more powerful the mirror.

The convex mirror shown in [\[link\]](#) also has a focal point. Parallel rays of light reflected from the mirror seem to originate from the point F at the

focal distance f behind the mirror. The focal length and power of a convex mirror are negative, since it is a diverging mirror.



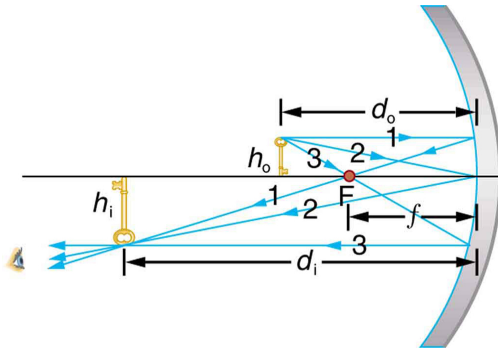
Parallel rays of light reflected from a convex spherical mirror (small in size compared with its radius of curvature) seem to originate from a well-defined focal point at the focal distance f behind the mirror. Convex mirrors diverge light rays and, thus, have a negative focal length.

Ray tracing is as useful for mirrors as for lenses. The rules for ray tracing for mirrors are based on the illustrations just discussed:

1. A ray approaching a concave converging mirror parallel to its axis is reflected through the focal point F of the mirror on the same side. (See rays 1 and 3 in [\[link\]](#)(b).)
2. A ray approaching a convex diverging mirror parallel to its axis is reflected so that it seems to come from the focal point F behind the mirror. (See rays 1 and 3 in [\[link\]](#).)
3. Any ray striking the center of a mirror is followed by applying the law of reflection; it makes the same angle with the axis when leaving as when approaching. (See ray 2 in [\[link\]](#).)
4. A ray approaching a concave converging mirror through its focal point is reflected parallel to its axis. (The reverse of rays 1 and 3 in [\[link\]](#).)
5. A ray approaching a convex diverging mirror by heading toward its focal point on the opposite side is reflected parallel to the axis. (The reverse of rays 1 and 3 in [\[link\]](#).)

We will use ray tracing to illustrate how images are formed by mirrors, and we can use ray tracing quantitatively to obtain numerical information. But since we assume each mirror is small compared with its radius of curvature, we can use the thin lens equations for mirrors just as we did for lenses.

Consider the situation shown in [\[link\]](#), concave spherical mirror reflection, in which an object is placed farther from a concave (converging) mirror than its focal length. That is, f is positive and $d_o > f$, so that we may expect an image similar to the case 1 real image formed by a converging lens. Ray tracing in [\[link\]](#) shows that the rays from a common point on the object all cross at a point on the same side of the mirror as the object. Thus a real image can be projected onto a screen placed at this location. The image distance is positive, and the image is inverted, so its magnification is negative. This is a *case 1 image for mirrors*. It differs from the case 1 image for lenses only in that the image is on the same side of the mirror as the object. It is otherwise identical.



A case 1 image for a mirror. An object is farther from the converging mirror than its focal length. Rays from a common point on the object are traced using the rules in the text. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 goes through the focal point on the way toward the mirror. All three rays cross at the same point after being reflected, locating the inverted real image. Although three rays are shown, only two of the three are needed to locate the image and determine its height.

Example:

A Concave Reflector

Electric room heaters use a concave mirror to reflect infrared (IR) radiation from hot coils. Note that IR follows the same law of reflection as visible light. Given that the mirror has a radius of curvature of 50.0 cm and produces an image of the coils 3.00 m away from the mirror, where are the coils?

Strategy and Concept

We are given that the concave mirror projects a real image of the coils at an image distance $d_i = 3.00$ m. The coils are the object, and we are asked to find their location—that is, to find the object distance d_o . We are also given the radius of curvature of the mirror, so that its focal length is $f = R/2 = 25.0$ cm (positive since the mirror is concave or converging). Assuming the mirror is small compared with its radius of curvature, we can use the thin lens equations, to solve this problem.

Solution

Since d_i and f are known, thin lens equation can be used to find d_o :

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}.$$

Rearranging to isolate d_o gives

Equation:

$$\frac{1}{d_o} = \frac{1}{f} - \frac{1}{d_i}.$$

Entering known quantities gives a value for $1/d_o$:

Equation:

$$\frac{1}{d_o} = \frac{1}{0.250 \text{ m}} - \frac{1}{3.00 \text{ m}} = \frac{3.667}{\text{m}}.$$

This must be inverted to find d_o :

Equation:

$$d_o = \frac{1 \text{ m}}{3.667} = 27.3 \text{ cm}.$$

Discussion

Note that the object (the filament) is farther from the mirror than the mirror's focal length. This is a case 1 image ($d_o > f$ and f positive), consistent with the fact that a real image is formed. You will get the most concentrated thermal energy directly in front of the mirror and 3.00 m away from it. Generally, this is not desirable, since it could cause burns. Usually, you want the rays to emerge parallel, and this is accomplished by having the filament at the focal point of the mirror.

Note that the filament here is not much farther from the mirror than its focal length and that the image produced is considerably farther away. This is exactly analogous to a slide projector. Placing a slide only slightly farther away from the projector lens than its focal length produces an image significantly farther away. As the object gets closer to the focal distance, the image gets farther away. In fact, as the object distance approaches the focal length, the image distance approaches infinity and the rays are sent out parallel to one another.

Example:

Solar Electric Generating System

One of the solar technologies used today for generating electricity is a device (called a parabolic trough or concentrating collector) that concentrates the sunlight onto a blackened pipe that contains a fluid. This heated fluid is pumped to a heat exchanger, where its heat energy is transferred to another system that is used to generate steam—and so generate electricity through a conventional steam cycle. [\[link\]](#) shows such a working system in southern California. Concave mirrors are used to concentrate the sunlight onto the pipe. The mirror has the approximate shape of a section of a cylinder. For the problem, assume that the mirror is exactly one-quarter of a full cylinder.

- a. If we wish to place the fluid-carrying pipe 40.0 cm from the concave mirror at the mirror's focal point, what will be the radius of curvature of the mirror?
- b. Per meter of pipe, what will be the amount of sunlight concentrated onto the pipe, assuming the insolation (incident solar radiation) is

0.900 kW/m²?

- c. If the fluid-carrying pipe has a 2.00-cm diameter, what will be the temperature increase of the fluid per meter of pipe over a period of one minute? Assume all the solar radiation incident on the reflector is absorbed by the pipe, and that the fluid is mineral oil.

Strategy

To solve an *Integrated Concept Problem* we must first identify the physical principles involved. Part (a) is related to the current topic. Part (b) involves a little math, primarily geometry. Part (c) requires an understanding of heat and density.

Solution to (a)

To a good approximation for a concave or semi-spherical surface, the point where the parallel rays from the sun converge will be at the focal point, so $R = 2f = 80.0$ cm.

Solution to (b)

The insolation is 900 W/m². We must find the cross-sectional area A of the concave mirror, since the power delivered is $900 \text{ W/m}^2 \times A$. The mirror in this case is a quarter-section of a cylinder, so the area for a length L of the mirror is $A = \frac{1}{4}(2\pi R)L$. The area for a length of 1.00 m is then

Equation:

$$A = \frac{\pi}{2} R(1.00 \text{ m}) = \frac{(3.14)}{2} (0.800 \text{ m})(1.00 \text{ m}) = 1.26 \text{ m}^2.$$

The insolation on the 1.00-m length of pipe is then

Equation:

$$\left(9.00 \times 10^2 \frac{\text{W}}{\text{m}^2}\right) \left(1.26 \text{ m}^2\right) = 1130 \text{ W}.$$

Solution to (c)

The increase in temperature is given by $Q = mc \Delta T$. The mass m of the mineral oil in the one-meter section of pipe is

Equation:

$$\begin{aligned}
 m &= \rho V = \rho \pi \left(\frac{d}{2} \right)^2 (1.00 \text{ m}) \\
 &= (8.00 \times 10^2 \text{ kg/m}^3)(3.14)(0.0100 \text{ m})^2(1.00 \text{ m}) \\
 &= 0.251 \text{ kg}.
 \end{aligned}$$

Therefore, the increase in temperature in one minute is

Equation:

$$\begin{aligned}
 \Delta T &= Q/mc \\
 &= \frac{(1130 \text{ W})(60.0 \text{ s})}{(0.251 \text{ kg})(1670 \text{ J}\cdot\text{kg}/^\circ\text{C})} \\
 &= 162^\circ\text{C}.
 \end{aligned}$$

Discussion for (c)

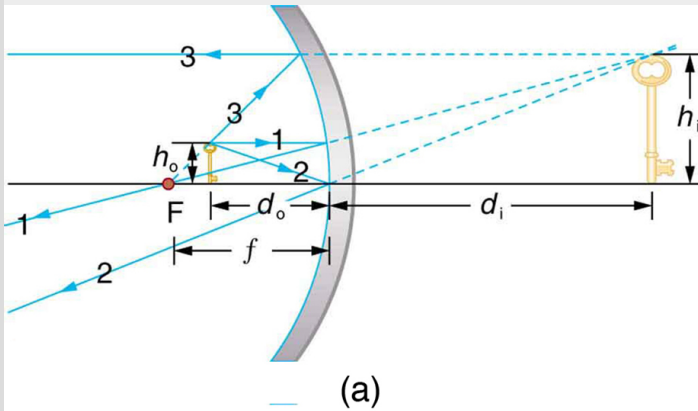
An array of such pipes in the California desert can provide a thermal output of 250 MW on a sunny day, with fluids reaching temperatures as high as 400°C. We are considering only one meter of pipe here, and ignoring heat losses along the pipe.



Parabolic trough collectors are used to generate electricity in southern California. (credit: kjkolb, Wikimedia Commons)

What happens if an object is closer to a concave mirror than its focal length? This is analogous to a case 2 image for lenses ($d_o < f$ and f positive), which is a magnifier. In fact, this is how makeup mirrors act as magnifiers. [\[link\]](#)(a) uses ray tracing to locate the image of an object placed close to a concave mirror. Rays from a common point on the object

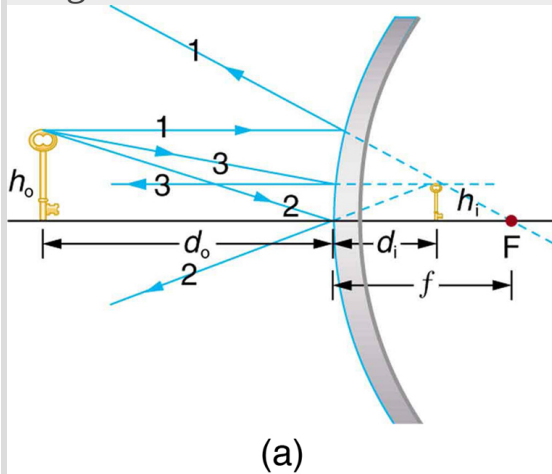
are reflected in such a manner that they appear to be coming from behind the mirror, meaning that the image is virtual and cannot be projected. As with a magnifying glass, the image is upright and larger than the object. This is a *case 2 image for mirrors* and is exactly analogous to that for lenses.



(a) Case 2 images for mirrors are formed when a converging mirror has an object closer to it than its focal length. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 approaches the mirror as if it came from the focal point. (b) A magnifying mirror showing the reflection. (credit: Mike Melrose, Flickr)

All three rays appear to originate from the same point after being reflected, locating the upright virtual image behind the mirror and showing it to be larger than the object. (b) Makeup mirrors are perhaps the most common use of a concave mirror to produce a larger, upright image.

A convex mirror is a diverging mirror (f is negative) and forms only one type of image. It is a *case 3* image—one that is upright and smaller than the object, just as for diverging lenses. [\[link\]](#)(a) uses ray tracing to illustrate the location and size of the case 3 image for mirrors. Since the image is behind the mirror, it cannot be projected and is thus a virtual image. It is also seen to be smaller than the object.



Case 3 images for mirrors are formed by any convex mirror. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the

mirror, and ray 3 approaches toward the focal point. All three rays appear to originate from the same point after being reflected, locating the upright virtual image behind the mirror and showing it to be smaller than the object.

(b) Security mirrors are convex, producing a smaller, upright image. Because the image is smaller, a larger area is imaged compared to what would be observed for a flat mirror (and hence security is improved).

(credit: Laura D'Alessandro, Flickr)

Example:

Image in a Convex Mirror

A keratometer is a device used to measure the curvature of the cornea, particularly for fitting contact lenses. Light is reflected from the cornea, which acts like a convex mirror, and the keratometer measures the magnification of the image. The smaller the magnification, the smaller the radius of curvature of the cornea. If the light source is 12.0 cm from the cornea and the image's magnification is 0.0320, what is the cornea's radius of curvature?

Strategy

If we can find the focal length of the convex mirror formed by the cornea, we can find its radius of curvature (the radius of curvature is twice the focal length of a spherical mirror). We are given that the object distance is

$d_o = 12.0$ cm and that $m = 0.0320$. We first solve for the image distance d_i , and then for f .

Solution

$m = -d_i/d_o$. Solving this expression for d_i gives

Equation:

$$d_i = -md_o.$$

Entering known values yields

Equation:

$$d_i = -(0.0320)(12.0 \text{ cm}) = -0.384 \text{ cm}.$$

Equation:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$$

Substituting known values,

Equation:

$$\frac{1}{f} = \frac{1}{12.0 \text{ cm}} + \frac{1}{-0.384 \text{ cm}} = \frac{-2.52}{\text{cm}}.$$

This must be inverted to find f :

Equation:

$$f = \frac{\text{cm}}{-2.52} = -0.400 \text{ cm}.$$

The radius of curvature is twice the focal length, so that

Equation:

$$R = 2 | f | = 0.800 \text{ cm}.$$

Discussion

Although the focal length f of a convex mirror is defined to be negative, we take the absolute value to give us a positive value for R . The radius of curvature found here is reasonable for a cornea. The distance from cornea

to retina in an adult eye is about 2.0 cm. In practice, many corneas are not spherical, complicating the job of fitting contact lenses. Note that the image distance here is negative, consistent with the fact that the image is behind the mirror, where it cannot be projected. In this section's Problems and Exercises, you will show that for a fixed object distance, the smaller the radius of curvature, the smaller the magnification.

The three types of images formed by mirrors (cases 1, 2, and 3) are exactly analogous to those formed by lenses, as summarized in the table at the end of [Image Formation by Lenses](#). It is easiest to concentrate on only three types of images—then remember that concave mirrors act like convex lenses, whereas convex mirrors act like concave lenses.

Note:

Take-Home Experiment: Concave Mirrors Close to Home

Find a flashlight and identify the curved mirror used in it. Find another flashlight and shine the first flashlight onto the second one, which is turned off. Estimate the focal length of the mirror. You might try shining a flashlight on the curved mirror behind the headlight of a car, keeping the headlight switched off, and determine its focal length.

Problem-Solving Strategy for Mirrors

Step 1. Examine the situation to determine that image formation by a mirror is involved.

Step 2. Refer to the [Problem-Solving Strategies for Lenses](#). The same strategies are valid for mirrors as for lenses with one qualification—use the ray tracing rules for mirrors listed earlier in this section.

Section Summary

- The characteristics of an image formed by a flat mirror are: (a) The image and object are the same distance from the mirror, (b) The image

- is a virtual image, and (c) The image is situated behind the mirror.
- Image length is half the radius of curvature.

Equation:

$$f = \frac{R}{2}$$

- A convex mirror is a diverging mirror and forms only one type of image, namely a virtual image.

Conceptual Questions

Exercise:

Problem:

What are the differences between real and virtual images? How can you tell (by looking) whether an image formed by a single lens or mirror is real or virtual?

Exercise:

Problem:

Can you see a virtual image? Can you photograph one? Can one be projected onto a screen with additional lenses or mirrors? Explain your responses.

Exercise:

Problem:

Is it necessary to project a real image onto a screen for it to exist?

Exercise:

Problem:

At what distance is an image *always* located—at d_o , d_i , or f ?

Exercise:

Problem:

Under what circumstances will an image be located at the focal point of a lens or mirror?

Exercise:**Problem:**

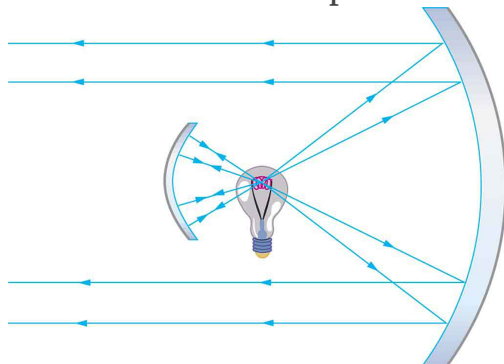
What is meant by a negative magnification? What is meant by a magnification that is less than 1 in magnitude?

Exercise:**Problem:**

Can a case 1 image be larger than the object even though its magnification is always negative? Explain.

Exercise:**Problem:**

[\[link\]](#) shows a light bulb between two mirrors. One mirror produces a beam of light with parallel rays; the other keeps light from escaping without being put into the beam. Where is the filament of the light in relation to the focal point or radius of curvature of each mirror?



The two mirrors trap most of the bulb's light and form a directional beam as in a headlight.

Exercise:**Problem:**

Devise an arrangement of mirrors allowing you to see the back of your head. What is the minimum number of mirrors needed for this task?

Exercise:**Problem:**

If you wish to see your entire body in a flat mirror (from head to toe), how tall should the mirror be? Does its size depend upon your distance away from the mirror? Provide a sketch.

Exercise:**Problem:**

It can be argued that a flat mirror has an infinite focal length. If so, where does it form an image? That is, how are d_i and d_o related?

Exercise:**Problem:**

Why are diverging mirrors often used for rear-view mirrors in vehicles? What is the main disadvantage of using such a mirror compared with a flat one?

Problems & Exercises**Exercise:****Problem:**

What is the focal length of a makeup mirror that has a power of 1.50 D?

Solution:

+0.667 m

Exercise:

Problem:

Some telephoto cameras use a mirror rather than a lens. What radius of curvature mirror is needed to replace a 800 mm focal length telephoto lens?

Exercise:

Problem:

(a) Calculate the focal length of the mirror formed by the shiny back of a spoon that has a 3.00 cm radius of curvature. (b) What is its power in diopters?

Solution:

(a) -1.5×10^{-2} m

(b) -66.7 D

Exercise:

Problem:

Find the magnification of the heater element in [\[link\]](#). Note that its large magnitude helps spread out the reflected energy.

Exercise:

Problem:

What is the focal length of a makeup mirror that produces a magnification of 1.50 when a person's face is 12.0 cm away? Explicitly show how you follow the steps in the [Problem-Solving Strategy for Mirrors](#).

Solution:

+0.360 m (concave)

Exercise:

Problem:

A shopper standing 3.00 m from a convex security mirror sees his image with a magnification of 0.250. (a) Where is his image? (b) What is the focal length of the mirror? (c) What is its radius of curvature? Explicitly show how you follow the steps in the [Problem-Solving Strategy for Mirrors](#).

Exercise:

Problem:

An object 1.50 cm high is held 3.00 cm from a person's cornea, and its reflected image is measured to be 0.167 cm high. (a) What is the magnification? (b) Where is the image? (c) Find the radius of curvature of the convex mirror formed by the cornea. (Note that this technique is used by optometrists to measure the curvature of the cornea for contact lens fitting. The instrument used is called a keratometer, or curve measurer.)

Solution:

(a) +0.111

(b) -0.334 cm (behind "mirror")

(c) 0.752cm

Exercise:

Problem:

Ray tracing for a flat mirror shows that the image is located a distance behind the mirror equal to the distance of the object from the mirror. This is stated $d_i = -d_o$, since this is a negative image distance (it is a virtual image). (a) What is the focal length of a flat mirror? (b) What is its power?

Exercise:**Problem:**

Show that for a flat mirror $h_i = h_o$, knowing that the image is a distance behind the mirror equal in magnitude to the distance of the object from the mirror.

Solution:**Equation:**

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o} = -\frac{-d_o}{d_o} = \frac{d_o}{d_o} = 1 \Rightarrow h_i = h_o$$

Exercise:**Problem:**

Use the law of reflection to prove that the focal length of a mirror is half its radius of curvature. That is, prove that $f = R/2$. Note this is true for a spherical mirror only if its diameter is small compared with its radius of curvature.

Exercise:**Problem:**

Referring to the electric room heater considered in the first example in this section, calculate the intensity of IR radiation in W/m^2 projected by the concave mirror on a person 3.00 m away. Assume that the heating element radiates 1500 W and has an area of 100 cm^2 , and that half of the radiated power is reflected and focused by the mirror.

Solution:

$$6.82 \text{ kW}/\text{m}^2$$

Exercise:

Problem:

Consider a 250-W heat lamp fixed to the ceiling in a bathroom. If the filament in one light burns out then the remaining three still work. Construct a problem in which you determine the resistance of each filament in order to obtain a certain intensity projected on the bathroom floor. The ceiling is 3.0 m high. The problem will need to involve concave mirrors behind the filaments. Your instructor may wish to guide you on the level of complexity to consider in the electrical components.

Glossary

converging mirror

a concave mirror in which light rays that strike it parallel to its axis converge at one or more points along the axis

diverging mirror

a convex mirror in which light rays that strike it parallel to its axis bend away (diverge) from its axis

law of reflection

angle of reflection equals the angle of incidence

Introduction to Vision and Optical Instruments

class="introduction"

A scientist
examines
minute
details on the
surface of a
disk drive at
a
magnification
of 100,000
times. The
image was
produced
using an
electron
microscope.
(credit:
Robert
Scoble)



Explore how the image on the computer screen is formed. How is the image formation on the computer screen different from the image formation in your eye as you look down the microscope? How can videos of living cell processes be taken for viewing later on, and by many different people?

Seeing faces and objects we love and cherish is a delight—one's favorite teddy bear, a picture on the wall, or the sun rising over the mountains. Intricate images help us understand nature and are invaluable for developing techniques and technologies in order to improve the quality of life. The image of a red blood cell that almost fills the cross-sectional area of a tiny capillary makes us wonder how blood makes it through and not get stuck. We are able to see bacteria and viruses and understand their structure. It is the knowledge of physics that provides fundamental understanding and models required to develop new techniques and instruments. Therefore, physics is called an *enabling science*—a science that enables development and advancement in other areas. It is through optics and imaging that physics enables advancement in major areas of biosciences. This chapter illustrates the enabling nature of physics through an understanding of how a

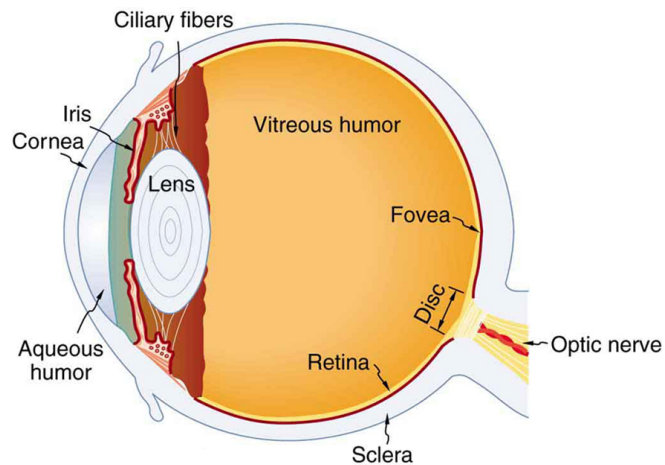
human eye is able to see and how we are able to use optical instruments to see beyond what is possible with the naked eye. It is convenient to categorize these instruments on the basis of geometric optics (see [Geometric Optics](#)) and wave optics (see [Wave Optics](#)).

Physics of the Eye

- Explain the image formation by the eye.
- Explain why peripheral images lack detail and color.
- Define refractive indices.
- Analyze the accommodation of the eye for distant and near vision.

The eye is perhaps the most interesting of all optical instruments. The eye is remarkable in how it forms images and in the richness of detail and color it can detect. However, our eyes commonly need some correction, to reach what is called “normal” vision, but should be called ideal rather than normal. Image formation by our eyes and common vision correction are easy to analyze with the optics discussed in [Geometric Optics](#).

[\[link\]](#) shows the basic anatomy of the eye. The cornea and lens form a system that, to a good approximation, acts as a single thin lens. For clear vision, a real image must be projected onto the light-sensitive retina, which lies at a fixed distance from the lens. The lens of the eye adjusts its power to produce an image on the retina for objects at different distances. The center of the image falls on the fovea, which has the greatest density of light receptors and the greatest acuity (sharpness) in the visual field. The variable opening (or pupil) of the eye along with chemical adaptation allows the eye to detect light intensities from the lowest observable to 10^{10} times greater (without damage). This is an incredible range of detection. Our eyes perform a vast number of functions, such as sense direction, movement, sophisticated colors, and distance. Processing of visual nerve impulses begins with interconnections in the retina and continues in the brain. The optic nerve conveys signals received by the eye to the brain.



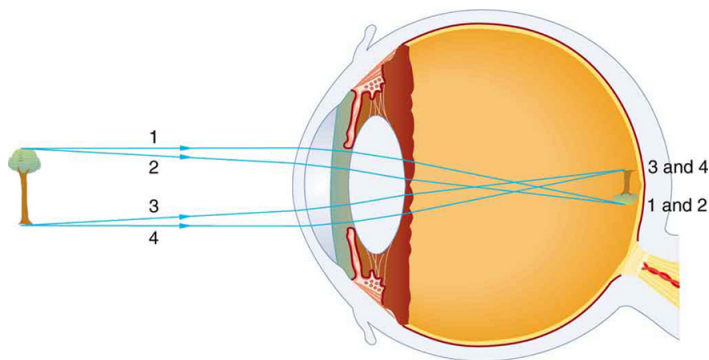
The cornea and lens of an eye act together to form a real image on the light-sensing retina, which has its densest concentration of receptors in the fovea and a blind spot over the optic nerve. The power of the lens of an eye is adjustable to provide an image on the retina for varying object distances. Layers of tissues with varying indices of refraction in the lens are shown here. However, they have been omitted from other pictures for clarity.

Refractive indices are crucial to image formation using lenses. [\[link\]](#) shows refractive indices relevant to the eye. The biggest change in the refractive index, and bending of rays, occurs at the cornea rather than the lens. The ray diagram in [\[link\]](#) shows image formation by the cornea and lens of the eye. The rays bend according to the refractive indices provided in [\[link\]](#). The cornea provides about two-thirds of the power of the eye, owing to the fact that speed of light changes considerably while traveling from air into cornea. The lens provides the remaining power needed to produce an image on the retina. The cornea and lens can be treated as a single thin lens, even

though the light rays pass through several layers of material (such as cornea, aqueous humor, several layers in the lens, and vitreous humor), changing direction at each interface. The image formed is much like the one produced by a single convex lens. This is a case 1 image. Images formed in the eye are inverted but the brain inverts them once more to make them seem upright.

Material	Index of Refraction
Water	1.33
Air	1.0
Cornea	1.38
Aqueous humor	1.34
Lens	1.41 average (varies throughout the lens, greatest in center)
Vitreous humor	1.34

Refractive Indices Relevant to the Eye

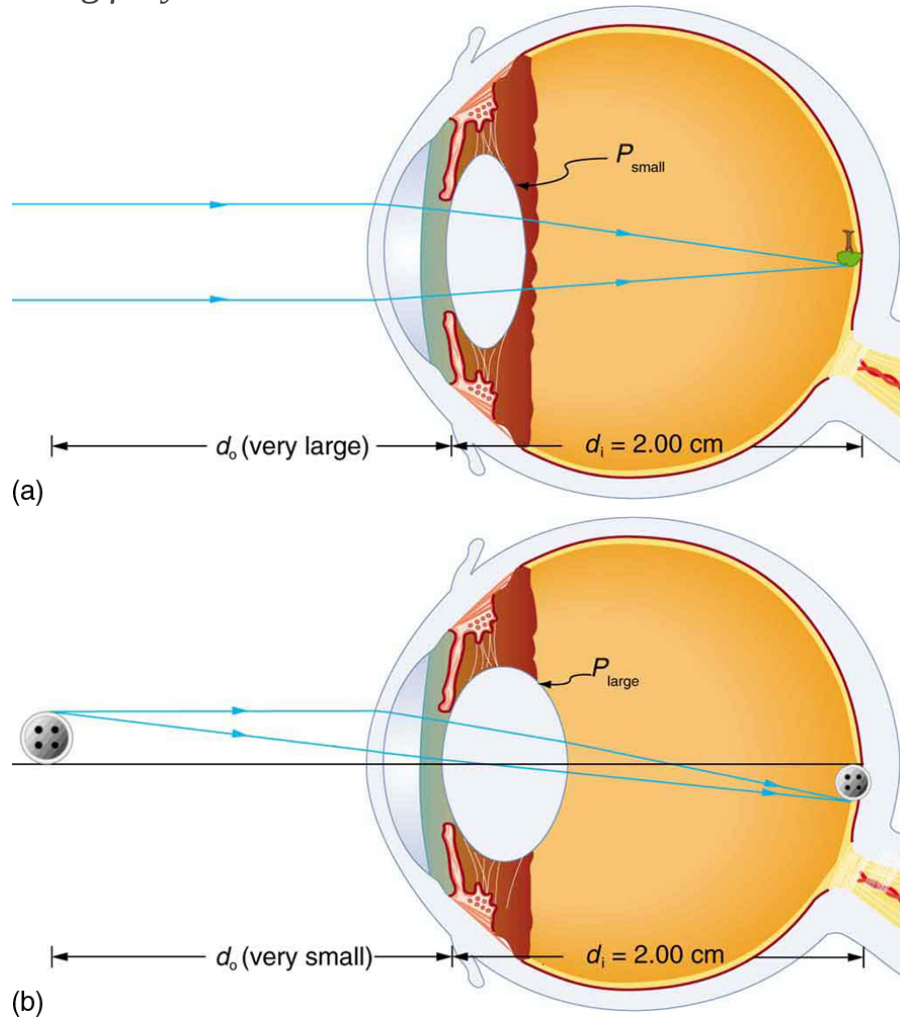


An image is formed on the retina with light rays converging most at the cornea and upon entering and exiting the lens. Rays from the top and bottom of the object are traced and produce an inverted real image on the retina. The distance to the object is drawn smaller than scale.

As noted, the image must fall precisely on the retina to produce clear vision — that is, the image distance d_i must equal the lens-to-retina distance. Because the lens-to-retina distance does not change, the image distance d_i must be the same for objects at all distances. The eye manages this by varying the power (and focal length) of the lens to accommodate for objects at various distances. The process of adjusting the eye's focal length is called **accommodation**. A person with normal (ideal) vision can see objects clearly at distances ranging from 25 cm to essentially infinity. However, although the near point (the shortest distance at which a sharp focus can be obtained) increases with age (becoming meters for some older people), we will consider it to be 25 cm in our treatment here.

[\[link\]](#) shows the accommodation of the eye for distant and near vision. Since light rays from a nearby object can diverge and still enter the eye, the lens must be more converging (more powerful) for close vision than for distant vision. To be more converging, the lens is made thicker by the action of the ciliary muscle surrounding it. The eye is most relaxed when viewing

distant objects, one reason that microscopes and telescopes are designed to produce distant images. Vision of very distant objects is called *totally relaxed*, while close vision is termed *accommodated*, with the closest vision being *fully accommodated*.



Relaxed and accommodated vision for distant and close objects. (a) Light rays from the same point on a distant object must be nearly parallel while entering the eye and more easily converge to produce an image on the retina. (b) Light rays from a nearby object can diverge more and still enter the eye. A more powerful lens is needed to converge them on the retina than if they were parallel.

We will use the thin lens equations to examine image formation by the eye quantitatively. First, note the power of a lens is given as $p = 1/f$, so we rewrite the thin lens equations as

Equation:

$$P = \frac{1}{d_o} + \frac{1}{d_i}$$

and

Equation:

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m.$$

We understand that d_i must equal the lens-to-retina distance to obtain clear vision, and that normal vision is possible for objects at distances $d_o = 25$ cm to infinity.

Note:

Take-Home Experiment: The Pupil

Look at the central transparent area of someone's eye, the pupil, in normal room light. Estimate the diameter of the pupil. Now turn off the lights and darken the room. After a few minutes turn on the lights and promptly estimate the diameter of the pupil. What happens to the pupil as the eye adjusts to the room light? Explain your observations.

The eye can detect an impressive amount of detail, considering how small the image is on the retina. To get some idea of how small the image can be, consider the following example.

Example:**Size of Image on Retina**

What is the size of the image on the retina of a 1.20×10^{-2} cm diameter human hair, held at arm's length (60.0 cm) away? Take the lens-to-retina distance to be 2.00 cm.

Strategy

We want to find the height of the image h_i , given the height of the object is $h_o = 1.20 \times 10^{-2}$ cm. We also know that the object is 60.0 cm away, so that $d_o = 60.0$ cm. For clear vision, the image distance must equal the lens-to-retina distance, and so $d_i = 2.00$ cm. The equation

$\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m$ can be used to find h_i with the known information.

Solution

The only unknown variable in the equation $\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m$ is h_i :

Equation:

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o}.$$

Rearranging to isolate h_i yields

Equation:

$$h_i = -h_o \cdot \frac{d_i}{d_o}.$$

Substituting the known values gives

Equation:

$$\begin{aligned} h_i &= -(1.20 \times 10^{-2} \text{ cm}) \frac{2.00 \text{ cm}}{60.0 \text{ cm}} \\ &= -4.00 \times 10^{-4} \text{ cm.} \end{aligned}$$

Discussion

This truly small image is not the smallest discernible—that is, the limit to visual acuity is even smaller than this. Limitations on visual acuity have to do with the wave properties of light and will be discussed in the next chapter. Some limitation is also due to the inherent anatomy of the eye and processing that occurs in our brain.

Example:**Power Range of the Eye**

Calculate the power of the eye when viewing objects at the greatest and smallest distances possible with normal vision, assuming a lens-to-retina distance of 2.00 cm (a typical value).

Strategy

For clear vision, the image must be on the retina, and so $d_i = 2.00$ cm here. For distant vision, $d_o \approx \infty$, and for close vision, $d_o = 25.0$ cm, as discussed earlier. The equation $P = \frac{1}{d_o} + \frac{1}{d_i}$ as written just above, can be used directly to solve for P in both cases, since we know d_i and d_o . Power has units of diopters, where $1 \text{ D} = 1/\text{m}$, and so we should express all distances in meters.

Solution

For distant vision,

Equation:

$$P = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{\infty} + \frac{1}{0.0200 \text{ m}}.$$

Since $1/\infty = 0$, this gives

Equation:

$$P = 0 + 50.0/\text{m} = 50.0 \text{ D (distant vision)}.$$

Now, for close vision,

Equation:

$$\begin{aligned} P &= \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{0.250 \text{ m}} + \frac{1}{0.0200 \text{ m}} \\ &= \frac{4.00}{\text{m}} + \frac{50.0}{\text{m}} = 4.00 \text{ D} + 50.0 \text{ D} \\ &= 54.0 \text{ D (close vision)}. \end{aligned}$$

Discussion

For an eye with this typical 2.00 cm lens-to-retina distance, the power of the eye ranges from 50.0 D (for distant totally relaxed vision) to 54.0 D (for close fully accommodated vision), which is an 8% increase. This increase in power for close vision is consistent with the preceding

discussion and the ray tracing in [\[link\]](#). An 8% ability to accommodate is considered normal but is typical for people who are about 40 years old. Younger people have greater accommodation ability, whereas older people gradually lose the ability to accommodate. When an optometrist identifies accommodation as a problem in elder people, it is most likely due to stiffening of the lens. The lens of the eye changes with age in ways that tend to preserve the ability to see distant objects clearly but do not allow the eye to accommodate for close vision, a condition called **presbyopia** (literally, elder eye). To correct this vision defect, we place a converging, positive power lens in front of the eye, such as found in reading glasses. Commonly available reading glasses are rated by their power in diopters, typically ranging from 1.0 to 3.5 D.

Section Summary

- Image formation by the eye is adequately described by the thin lens equations:

Equation:

$$P = \frac{1}{d_o} + \frac{1}{d_i} \text{ and } \frac{h_i}{h_o} = -\frac{d_i}{d_o} = m.$$

- The eye produces a real image on the retina by adjusting its focal length and power in a process called accommodation.
- For close vision, the eye is fully accommodated and has its greatest power, whereas for distant vision, it is totally relaxed and has its smallest power.
- The loss of the ability to accommodate with age is called presbyopia, which is corrected by the use of a converging lens to add power for close vision.

Conceptual Questions

Exercise:

Problem:

If the lens of a person's eye is removed because of cataracts (as has been done since ancient times), why would you expect a spectacle lens of about 16 D to be prescribed?

Exercise:**Problem:**

A cataract is cloudiness in the lens of the eye. Is light dispersed or diffused by it?

Exercise:**Problem:**

When laser light is shone into a relaxed normal-vision eye to repair a tear by spot-welding the retina to the back of the eye, the rays entering the eye must be parallel. Why?

Exercise:**Problem:**

How does the power of a dry contact lens compare with its power when resting on the tear layer of the eye? Explain.

Exercise:**Problem:**

Why is your vision so blurry when you open your eyes while swimming under water? How does a face mask enable clear vision?

Problem Exercises

Unless otherwise stated, the lens-to-retina distance is 2.00 cm.

Exercise:

Problem:

What is the power of the eye when viewing an object 50.0 cm away?

Solution:

52.0 D

Exercise:**Problem:**

Calculate the power of the eye when viewing an object 3.00 m away.

Exercise:**Problem:**

(a) The print in many books averages 3.50 mm in height. How high is the image of the print on the retina when the book is held 30.0 cm from the eye?

(b) Compare the size of the print to the sizes of rods and cones in the fovea and discuss the possible details observable in the letters. (The eye-brain system can perform better because of interconnections and higher order image processing.)

Solution:

(a) -0.233 mm

(b) The size of the rods and the cones is smaller than the image height, so we can distinguish letters on a page.

Exercise:

Problem:

Suppose a certain person's visual acuity is such that he can see objects clearly that form an image $4.00\text{ }\mu\text{m}$ high on his retina. What is the maximum distance at which he can read the 75.0 cm high letters on the side of an airplane?

Exercise:**Problem:**

People who do very detailed work close up, such as jewellers, often can see objects clearly at much closer distance than the normal 25 cm .

- (a) What is the power of the eyes of a woman who can see an object clearly at a distance of only 8.00 cm ?
- (b) What is the size of an image of a 1.00 mm object, such as lettering inside a ring, held at this distance?
- (c) What would the size of the image be if the object were held at the normal 25.0 cm distance?

Solution:

- (a) $+62.5\text{ D}$
- (b) -0.250 mm
- (c) -0.0800 mm

Glossary

accommodation

the ability of the eye to adjust its focal length is known as accommodation

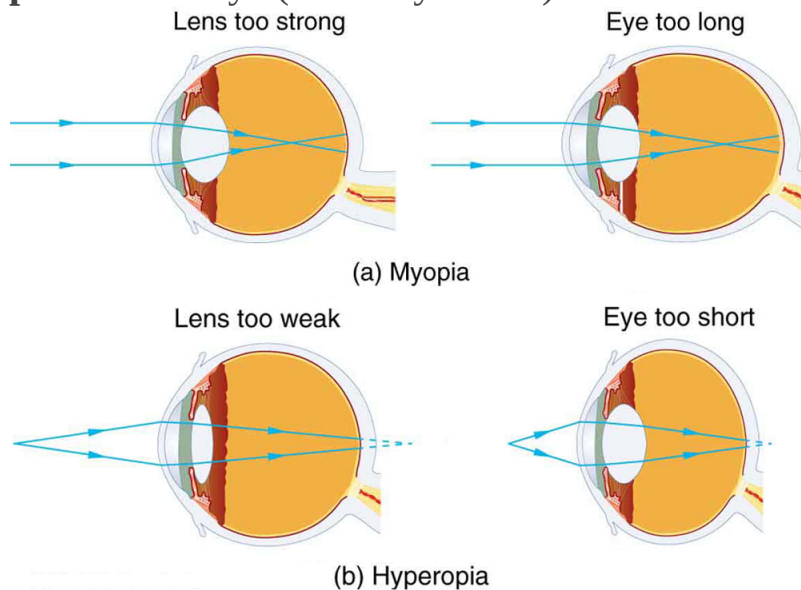
presbyopia

a condition in which the lens of the eye becomes progressively unable to focus on objects close to the viewer

Vision Correction

- Identify and discuss common vision defects.
- Explain nearsightedness and farsightedness corrections.
- Explain laser vision correction.

The need for some type of vision correction is very common. Common vision defects are easy to understand, and some are simple to correct. [\[link\]](#) illustrates two common vision defects. **Nearsightedness**, or **myopia**, is the inability to see distant objects clearly while close objects are clear. The eye overconverges the nearly parallel rays from a distant object, and the rays cross in front of the retina. More divergent rays from a close object are converged on the retina for a clear image. The distance to the farthest object that can be seen clearly is called the **far point** of the eye (normally infinity). **Farsightedness**, or **hyperopia**, is the inability to see close objects clearly while distant objects may be clear. A farsighted eye does not converge sufficient rays from a close object to make the rays meet on the retina. Less diverging rays from a distant object can be converged for a clear image. The distance to the closest object that can be seen clearly is called the **near point** of the eye (normally 25 cm).

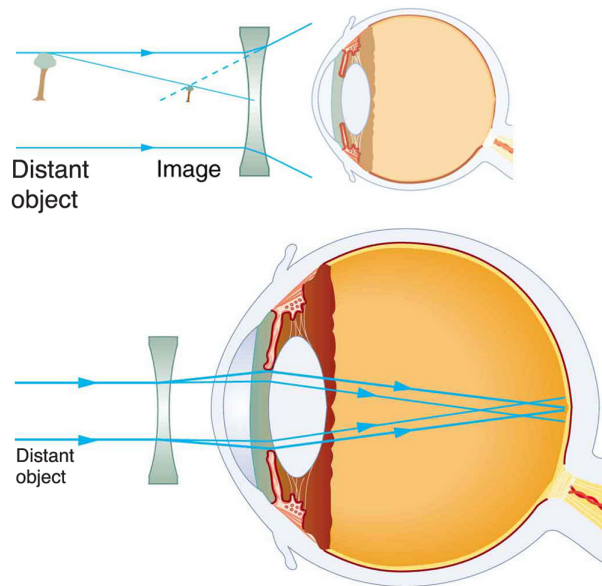


(a) The nearsighted (myopic) eye converges rays from a distant object in front of the retina; thus, they are diverging when they

strike the retina, producing a blurry image.

This can be caused by the lens of the eye being too powerful or the length of the eye being too great. (b) The farsighted (hyperopic) eye is unable to converge the rays from a close object by the time they strike the retina, producing blurry close vision. This can be caused by insufficient power in the lens or by the eye being too short.

Since the nearsighted eye over converges light rays, the correction for nearsightedness is to place a diverging spectacle lens in front of the eye. This reduces the power of an eye that is too powerful. Another way of thinking about this is that a diverging spectacle lens produces a case 3 image, which is closer to the eye than the object (see [\[link\]](#)). To determine the spectacle power needed for correction, you must know the person's far point—that is, you must know the greatest distance at which the person can see clearly. Then the image produced by a spectacle lens must be at this distance or closer for the nearsighted person to be able to see it clearly. It is worth noting that wearing glasses does not change the eye in any way. The eyeglass lens is simply used to create an image of the object at a distance where the nearsighted person can see it clearly. Whereas someone not wearing glasses can see clearly *objects* that fall between their near point and their far point, someone wearing glasses can see *images* that fall between their near point and their far point.



Correction of nearsightedness requires a diverging lens that compensates for the overconvergence by the eye. The diverging lens produces an image closer to the eye than the object, so that the nearsighted person can see it clearly.

Example:

Correcting Nearsightedness

What power of spectacle lens is needed to correct the vision of a nearsighted person whose far point is 30.0 cm? Assume the spectacle (corrective) lens is held 1.50 cm away from the eye by eyeglass frames.

Strategy

You want this nearsighted person to be able to see very distant objects clearly. That means the spectacle lens must produce an image 30.0 cm from the eye for an object very far away. An image 30.0 cm from the eye will be 28.5 cm to the left of the spectacle lens (see [\[link\]](#)). Therefore, we

must get $d_i = -28.5$ cm when $d_o \approx \infty$. The image distance is negative, because it is on the same side of the spectacle as the object.

Solution

Since d_i and d_o are known, the power of the spectacle lens can be found using $P = \frac{1}{d_o} + \frac{1}{d_i}$ as written earlier:

Equation:

$$P = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{\infty} + \frac{1}{-0.285 \text{ m}}.$$

Since $1/\infty = 0$, we obtain:

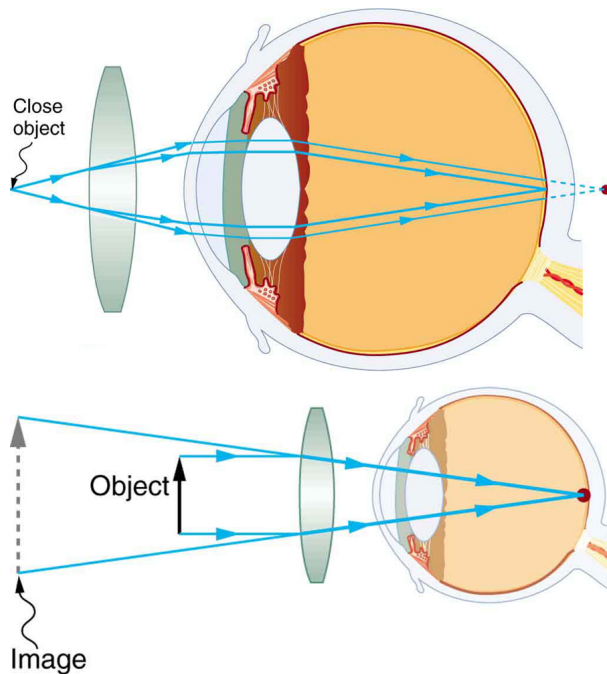
Equation:

$$P = 0 - 3.51/\text{m} = -3.51 \text{ D}.$$

Discussion

The negative power indicates a diverging (or concave) lens, as expected. The spectacle produces a case 3 image closer to the eye, where the person can see it. If you examine eyeglasses for nearsighted people, you will find the lenses are thinnest in the center. Additionally, if you examine a prescription for eyeglasses for nearsighted people, you will find that the prescribed power is negative and given in units of diopters.

Since the farsighted eye under converges light rays, the correction for farsightedness is to place a converging spectacle lens in front of the eye. This increases the power of an eye that is too weak. Another way of thinking about this is that a converging spectacle lens produces a case 2 image, which is farther from the eye than the object (see [\[link\]](#)). To determine the spectacle power needed for correction, you must know the person's near point—that is, you must know the smallest distance at which the person can see clearly. Then the image produced by a spectacle lens must be at this distance or farther for the farsighted person to be able to see it clearly.



Correction of farsightedness uses a converging lens that compensates for the under convergence by the eye. The converging lens produces an image farther from the eye than the object, so that the farsighted person can see it clearly.

Example:

Correcting Farsightedness

What power of spectacle lens is needed to allow a farsighted person, whose near point is 1.00 m, to see an object clearly that is 25.0 cm away? Assume the spectacle (corrective) lens is held 1.50 cm away from the eye by eyeglass frames.

Strategy

When an object is held 25.0 cm from the person's eyes, the spectacle lens must produce an image 1.00 m away (the near point). An image 1.00 m

from the eye will be 98.5 cm to the left of the spectacle lens because the spectacle lens is 1.50 cm from the eye (see [\[link\]](#)). Therefore, $d_i = -98.5$ cm. The image distance is negative, because it is on the same side of the spectacle as the object. The object is 23.5 cm to the left of the spectacle, so that $d_o = 23.5$ cm.

Solution

Since d_i and d_o are known, the power of the spectacle lens can be found using $P = \frac{1}{d_o} + \frac{1}{d_i}$:

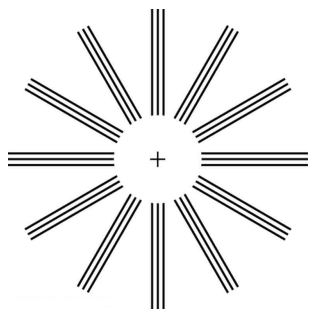
Equation:

$$\begin{aligned} P &= \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{0.235 \text{ m}} + \frac{1}{-0.985 \text{ m}} \\ &= 4.26 \text{ D} - 1.02 \text{ D} = 3.24 \text{ D}. \end{aligned}$$

Discussion

The positive power indicates a converging (convex) lens, as expected. The convex spectacle produces a case 2 image farther from the eye, where the person can see it. If you examine eyeglasses of farsighted people, you will find the lenses to be thickest in the center. In addition, a prescription of eyeglasses for farsighted people has a prescribed power that is positive.

Another common vision defect is **astigmatism**, an unevenness or asymmetry in the focus of the eye. For example, rays passing through a vertical region of the eye may focus closer than rays passing through a horizontal region, resulting in the image appearing elongated. This is mostly due to irregularities in the shape of the cornea but can also be due to lens irregularities or unevenness in the retina. Because of these irregularities, different parts of the lens system produce images at different locations. The eye-brain system can compensate for some of these irregularities, but they generally manifest themselves as less distinct vision or sharper images along certain axes. [\[link\]](#) shows a chart used to detect astigmatism. Astigmatism can be at least partially corrected with a spectacle having the opposite irregularity of the eye. If an eyeglass prescription has a cylindrical correction, it is there to correct astigmatism. The normal corrections for short- or farsightedness are spherical corrections, uniform along all axes.



This chart can detect astigmatism, unevenness in the focus of the eye. Check each of your eyes separately by looking at the center cross (without spectacles if you wear them). If lines along some axes appear darker or clearer than others, you have an astigmatism.

Contact lenses have advantages over glasses beyond their cosmetic aspects. One problem with glasses is that as the eye moves, it is not at a fixed distance from the spectacle lens. Contacts rest on and move with the eye, eliminating this problem. Because contacts cover a significant portion of the

cornea, they provide superior peripheral vision compared with eyeglasses. Contacts also correct some corneal astigmatism caused by surface irregularities. The tear layer between the smooth contact and the cornea fills in the irregularities. Since the index of refraction of the tear layer and the cornea are very similar, you now have a regular optical surface in place of an irregular one. If the curvature of a contact lens is not the same as the cornea (as may be necessary with some individuals to obtain a comfortable fit), the tear layer between the contact and cornea acts as a lens. If the tear layer is thinner in the center than at the edges, it has a negative power, for example. Skilled optometrists will adjust the power of the contact to compensate.

Laser vision correction has progressed rapidly in the last few years. It is the latest and by far the most successful in a series of procedures that correct vision by reshaping the cornea. As noted at the beginning of this section, the cornea accounts for about two-thirds of the power of the eye. Thus, small adjustments of its curvature have the same effect as putting a lens in front of the eye. To a reasonable approximation, the power of multiple lenses placed close together equals the sum of their powers. For example, a concave spectacle lens (for nearsightedness) having $P = -3.00$ D has the same effect on vision as reducing the power of the eye itself by 3.00 D. So to correct the eye for nearsightedness, the cornea is flattened to reduce its power. Similarly, to correct for farsightedness, the curvature of the cornea is enhanced to increase the power of the eye—the same effect as the positive power spectacle lens used for farsightedness. Laser vision correction uses high intensity electromagnetic radiation to ablate (to remove material from the surface) and reshape the corneal surfaces.

Today, the most commonly used laser vision correction procedure is *Laser in situ Keratomileusis (LASIK)*. The top layer of the cornea is surgically peeled back and the underlying tissue ablated by multiple bursts of finely controlled ultraviolet radiation produced by an excimer laser. Lasers are used because they not only produce well-focused intense light, but they also emit very pure wavelength electromagnetic radiation that can be controlled more accurately than mixed wavelength light. The 193 nm wavelength UV commonly used is extremely and strongly absorbed by corneal tissue,

allowing precise evaporation of very thin layers. A computer controlled program applies more bursts, usually at a rate of 10 per second, to the areas that require deeper removal. Typically a spot less than 1 mm in diameter and about 0.3 μm in thickness is removed by each burst. Nearsightedness, farsightedness, and astigmatism can be corrected with an accuracy that produces normal distant vision in more than 90% of the patients, in many cases right away. The corneal flap is replaced; healing takes place rapidly and is nearly painless. More than 1 million Americans per year undergo LASIK (see [\[link\]](#)).



Laser vision
correction is
being
performed
using the
LASIK
procedure.
Reshaping of
the cornea by
laser ablation is
based on a
careful
assessment of
the patient's
vision and is
computer
controlled. The

upper corneal
layer is
temporarily
peeled back
and minimally
disturbed in
LASIK,
providing for
more rapid and
less painful
healing of the
less sensitive
tissues below.
(credit: U.S.
Navy photo by
Mass
Communicatio
n Specialist 1st
Class Brien
Aho)

Section Summary

- Nearsightedness, or myopia, is the inability to see distant objects and is corrected with a diverging lens to reduce power.
- Farsightedness, or hyperopia, is the inability to see close objects and is corrected with a converging lens to increase power.
- In myopia and hyperopia, the corrective lenses produce images at a distance that the person can see clearly—the far point and near point, respectively.

Conceptual Questions

Exercise:

Problem:

It has become common to replace the cataract-clouded lens of the eye with an internal lens. This intraocular lens can be chosen so that the person has perfect distant vision. Will the person be able to read without glasses? If the person was nearsighted, is the power of the intraocular lens greater or less than the removed lens?

Exercise:**Problem:**

If the cornea is to be reshaped (this can be done surgically or with contact lenses) to correct myopia, should its curvature be made greater or smaller? Explain. Also explain how hyperopia can be corrected.

Exercise:**Problem:**

If there is a fixed percent uncertainty in LASIK reshaping of the cornea, why would you expect those people with the greatest correction to have a poorer chance of normal distant vision after the procedure?

Exercise:**Problem:**

A person with presbyopia has lost some or all of the ability to accommodate the power of the eye. If such a person's distant vision is corrected with LASIK, will she still need reading glasses? Explain.

Problem Exercises**Exercise:**

Problem:

What is the far point of a person whose eyes have a relaxed power of 50.5 D?

Solution:

2.00 m

Exercise:**Problem:**

What is the near point of a person whose eyes have an accommodated power of 53.5 D?

Exercise:**Problem:**

(a) A laser vision correction reshaping the cornea of a myopic patient reduces the power of his eye by 9.00 D, with a $\pm 5.0\%$ uncertainty in the final correction. What is the range of diopters for spectacle lenses that this person might need after LASIK procedure? (b) Was the person nearsighted or farsighted before the procedure? How do you know?

Solution:

(a) ± 0.45 D

(b) The person was nearsighted because the patient was myopic and the power was reduced.

Exercise:**Problem:**

In a LASIK vision correction, the power of a patient's eye is increased by 3.00 D. Assuming this produces normal close vision, what was the patient's near point before the procedure?

Exercise:**Problem:**

What was the previous far point of a patient who had laser vision correction that reduced the power of her eye by 7.00 D, producing normal distant vision for her?

Solution:

0.143 m

Exercise:**Problem:**

A severely myopic patient has a far point of 5.00 cm. By how many diopters should the power of his eye be reduced in laser vision correction to obtain normal distant vision for him?

Exercise:**Problem:**

A student's eyes, while reading the blackboard, have a power of 51.0 D. How far is the board from his eyes?

Solution:

1.00 m

Exercise:**Problem:**

The power of a physician's eyes is 53.0 D while examining a patient. How far from her eyes is the feature being examined?

Exercise:

Problem:

A young woman with normal distant vision has a 10.0% ability to accommodate (that is, increase) the power of her eyes. What is the closest object she can see clearly?

Solution:

20.0 cm

Exercise:**Problem:**

The far point of a myopic administrator is 50.0 cm. (a) What is the relaxed power of his eyes? (b) If he has the normal 8.00% ability to accommodate, what is the closest object he can see clearly?

Exercise:**Problem:**

A very myopic man has a far point of 20.0 cm. What power contact lens (when on the eye) will correct his distant vision?

Solution:

-5.00 D

Exercise:**Problem:**

Repeat the previous problem for eyeglasses held 1.50 cm from the eyes.

Exercise:**Problem:**

A myopic person sees that her contact lens prescription is -4.00 D. What is her far point?

Solution:

25.0 cm

Exercise:**Problem:**

Repeat the previous problem for glasses that are 1.75 cm from the eyes.

Exercise:**Problem:**

The contact lens prescription for a mildly farsighted person is 0.750 D, and the person has a near point of 29.0 cm. What is the power of the tear layer between the cornea and the lens if the correction is ideal, taking the tear layer into account?

Solution:

-0.198 D

Exercise:**Problem:**

A nearsighted man cannot see objects clearly beyond 20 cm from his eyes. How close must he stand to a mirror in order to see what he is doing when he shaves?

Exercise:**Problem:**

A mother sees that her child's contact lens prescription is 0.750 D. What is the child's near point?

Solution:

30.8 cm

Exercise:**Problem:**

Repeat the previous problem for glasses that are 2.20 cm from the eyes.

Exercise:**Problem:**

The contact lens prescription for a nearsighted person is -4.00 D and the person has a far point of 22.5 cm. What is the power of the tear layer between the cornea and the lens if the correction is ideal, taking the tear layer into account?

Solution:

-0.444 D

Exercise:**Problem: Unreasonable Results**

A boy has a near point of 50 cm and a far point of 500 cm. Will a -4.00 D lens correct his far point to infinity?

Glossary**nearsightedness**

another term for myopia, a visual defect in which distant objects appear blurred because their images are focused in front of the retina rather than being focused on the retina

myopia

a visual defect in which distant objects appear blurred because their images are focused in front of the retina rather than being focused on the retina

far point

the object point imaged by the eye onto the retina in an unaccommodated eye

farsightedness

another term for hyperopia, the condition of an eye where incoming rays of light reach the retina before they converge into a focused image

hyperopia

the condition of an eye where incoming rays of light reach the retina before they converge into a focused image

near point

the point nearest the eye at which an object is accurately focused on the retina at full accommodation

astigmatism

the result of an inability of the cornea to properly focus an image onto the retina

laser vision correction

a medical procedure used to correct astigmatism and eyesight deficiencies such as myopia and hyperopia

Color and Color Vision

- Explain the simple theory of color vision.
- Outline the coloring properties of light sources.
- Describe the retinex theory of color vision.

The gift of vision is made richer by the existence of color. Objects and lights abound with thousands of hues that stimulate our eyes, brains, and emotions. Two basic questions are addressed in this brief treatment—what does color mean in scientific terms, and how do we, as humans, perceive it?

Simple Theory of Color Vision

We have already noted that color is associated with the wavelength of visible electromagnetic radiation. When our eyes receive pure-wavelength light, we tend to see only a few colors. Six of these (most often listed) are red, orange, yellow, green, blue, and violet. These are the rainbow of colors produced when white light is dispersed according to different wavelengths. There are thousands of other **hues** that we can perceive. These include brown, teal, gold, pink, and white. One simple theory of color vision implies that all these hues are our eye's response to different combinations of wavelengths. This is true to an extent, but we find that color perception is even subtler than our eye's response for various wavelengths of light.

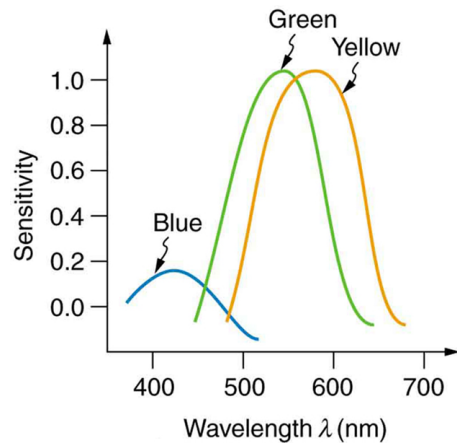
The two major types of light-sensing cells (photoreceptors) in the retina are **rods and cones**. Rods are more sensitive than cones by a factor of about 1000 and are solely responsible for peripheral vision as well as vision in very dark environments. They are also important for motion detection. There are about 120 million rods in the human retina. Rods do not yield color information. You may notice that you lose color vision when it is very dark, but you retain the ability to discern grey scales.

Note:

Take-Home Experiment: Rods and Cones

1. Go into a darkened room from a brightly lit room, or from outside in the Sun. How long did it take to start seeing shapes more clearly? What about color? Return to the bright room. Did it take a few minutes before you could see things clearly?
2. Demonstrate the sensitivity of foveal vision. Look at the letter G in the word ROGERS. What about the clarity of the letters on either side of G?

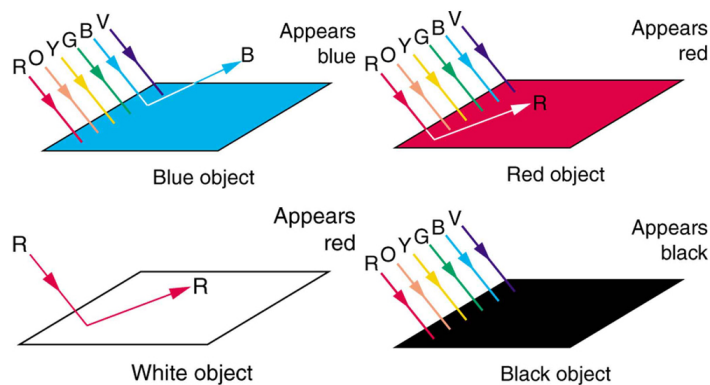
Cones are most concentrated in the fovea, the central region of the retina. There are no rods here. The fovea is at the center of the macula, a 5 mm diameter region responsible for our central vision. The cones work best in bright light and are responsible for high resolution vision. There are about 6 million cones in the human retina. There are three types of cones, and each type is sensitive to different ranges of wavelengths, as illustrated in [\[link\]](#). A **simplified theory of color vision** is that there are three *primary colors* corresponding to the three types of cones. The thousands of other hues that we can distinguish among are created by various combinations of stimulations of the three types of cones. Color television uses a three-color system in which the screen is covered with equal numbers of red, green, and blue phosphor dots. The broad range of hues a viewer sees is produced by various combinations of these three colors. For example, you will perceive yellow when red and green are illuminated with the correct ratio of intensities. White may be sensed when all three are illuminated. Then, it would seem that all hues can be produced by adding three primary colors in various proportions. But there is an indication that color vision is more sophisticated. There is no unique set of three primary colors. Another set that works is yellow, green, and blue. A further indication of the need for a more complex theory of color vision is that various different combinations can produce the same hue. Yellow can be sensed with yellow light, or with a combination of red and green, and also with white light from which violet has been removed. The three-primary-colors aspect of color vision is well established; more sophisticated theories expand on it rather than deny it.



The image shows the relative sensitivity of the three types of cones, which are named according to wavelengths of greatest sensitivity. Rods are about 1000 times more sensitive, and their curve peaks at about 500 nm. Evidence for the three types of cones comes from direct measurements in animal and human eyes and testing of color blind people.

Consider why various objects display color—that is, why are feathers blue and red in a crimson rosella? The *true color of an object* is defined by its absorptive or reflective characteristics. [\[link\]](#) shows white light falling on three different objects, one pure blue, one pure red, and one black, as well as pure red light falling on a white object. Other hues are created by more

complex absorption characteristics. Pink, for example on a galah cockatoo, can be due to weak absorption of all colors except red. An object can appear a different color under non-white illumination. For example, a pure blue object illuminated with pure red light will *appear* black, because it absorbs all the red light falling on it. But, the true color of the object is blue, which is independent of illumination.

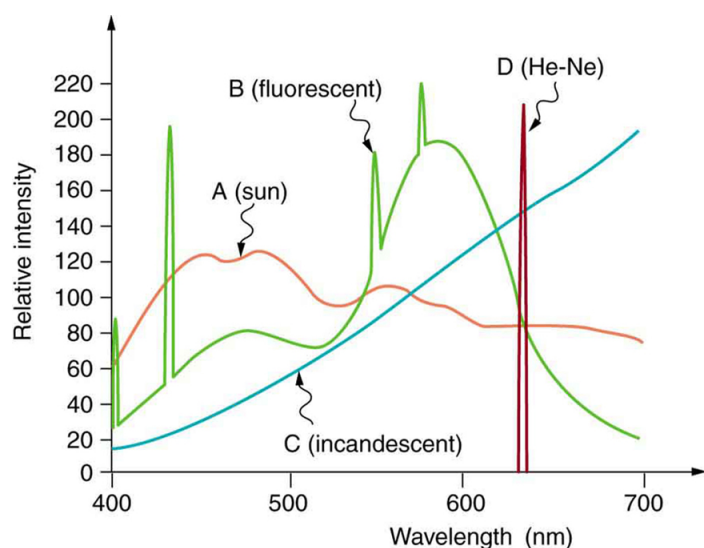


Absorption characteristics determine the true color of an object. Here, three objects are illuminated by white light, and one by pure red light. White is the equal mixture of all visible wavelengths; black is the absence of light.

Similarly, *light sources have colors* that are defined by the wavelengths they produce. A helium-neon laser emits pure red light. In fact, the phrase “pure red light” is defined by having a sharp constrained spectrum, a characteristic of laser light. The Sun produces a broad yellowish spectrum, fluorescent lights emit bluish-white light, and incandescent lights emit reddish-white hues as seen in [\[link\]](#). As you would expect, you sense these colors when viewing the light source directly or when illuminating a white object with them. All of this fits neatly into the simplified theory that a combination of wavelengths produces various hues.

Note:**Take-Home Experiment: Exploring Color Addition**

This activity is best done with plastic sheets of different colors as they allow more light to pass through to our eyes. However, thin sheets of paper and fabric can also be used. Overlay different colors of the material and hold them up to a white light. Using the theory described above, explain the colors you observe. You could also try mixing different crayon colors.

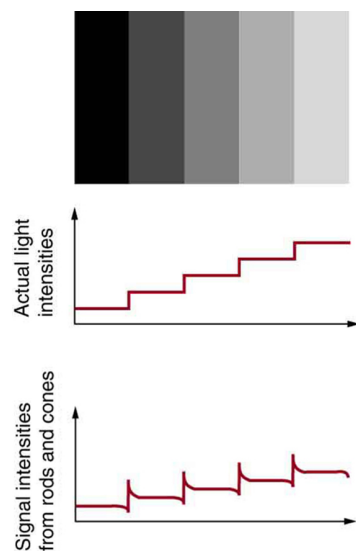


Emission spectra for various light sources are shown. Curve A is average sunlight at Earth's surface, curve B is light from a fluorescent lamp, and curve C is the output of an incandescent light. The spike for a helium-neon laser (curve D) is due to its pure wavelength emission. The spikes in the fluorescent output are due to atomic spectra—a topic that will be explored later.

Color Constancy and a Modified Theory of Color Vision

The eye-brain color-sensing system can, by comparing various objects in its view, perceive the true color of an object under varying lighting conditions—an ability that is called **color constancy**. We can sense that a white tablecloth, for example, is white whether it is illuminated by sunlight, fluorescent light, or candlelight. The wavelengths entering the eye are quite different in each case, as the graphs in [\[link\]](#) imply, but our color vision can detect the true color by comparing the tablecloth with its surroundings.

Theories that take color constancy into account are based on a large body of anatomical evidence as well as perceptual studies. There are nerve connections among the light receptors on the retina, and there are far fewer nerve connections to the brain than there are rods and cones. This means that there is signal processing in the eye before information is sent to the brain. For example, the eye makes comparisons between adjacent light receptors and is very sensitive to edges as seen in [\[link\]](#). Rather than responding simply to the light entering the eye, which is uniform in the various rectangles in this figure, the eye responds to the edges and senses false darkness variations.



The importance
of edges is

shown.
Although the
grey strips are
uniformly
shaded, as
indicated by the
graph
immediately
below them,
they do not
appear uniform
at all. Instead,
they are
perceived darker
on the dark side
and lighter on
the light side of
the edge, as
shown in the
bottom graph.
This is due to
nerve impulse
processing in
the eye.

One theory that takes various factors into account was advanced by Edwin Land (1909 – 1991), the creative founder of the Polaroid Corporation. Land proposed, based partly on his many elegant experiments, that the three types of cones are organized into systems called **retinexes**. Each retinex forms an image that is compared with the others, and the eye-brain system thus can compare a candle-illuminated white table cloth with its generally reddish surroundings and determine that it is actually white. This **retinex theory of color vision** is an example of modified theories of color vision that attempt to account for its subtleties. One striking experiment performed by Land demonstrates that some type of image comparison may produce color

vision. Two pictures are taken of a scene on black-and-white film, one using a red filter, the other a blue filter. Resulting black-and-white slides are then projected and superimposed on a screen, producing a black-and-white image, as expected. Then a red filter is placed in front of the slide taken with a red filter, and the images are again superimposed on a screen. You would expect an image in various shades of pink, but instead, the image appears to humans in full color with all the hues of the original scene. This implies that color vision can be induced by comparison of the black-and-white and red images. Color vision is not completely understood or explained, and the retinex theory is not totally accepted. It is apparent that color vision is much subtler than what a first look might imply.

Note:

PhET Explorations: Color Vision

Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

https://phet.colorado.edu/sims/html/color-vision/latest/color-vision_en.html

Section Summary

- The eye has four types of light receptors—rods and three types of color-sensitive cones.
- The rods are good for night vision, peripheral vision, and motion changes, while the cones are responsible for central vision and color.
- We perceive many hues, from light having mixtures of wavelengths.
- A simplified theory of color vision states that there are three primary colors, which correspond to the three types of cones, and that various combinations of the primary colors produce all the hues.
- The true color of an object is related to its relative absorption of various wavelengths of light. The color of a light source is related to the wavelengths it produces.

- Color constancy is the ability of the eye-brain system to discern the true color of an object illuminated by various light sources.
- The retinex theory of color vision explains color constancy by postulating the existence of three retinexes or image systems, associated with the three types of cones that are compared to obtain sophisticated information.

Conceptual Questions

Exercise:

Problem:

A pure red object on a black background seems to disappear when illuminated with pure green light. Explain why.

Exercise:

Problem: What is color constancy, and what are its limitations?

Exercise:

Problem:

There are different types of color blindness related to the malfunction of different types of cones. Why would it be particularly useful to study those rare individuals who are color blind only in one eye or who have a different type of color blindness in each eye?

Exercise:

Problem:

Propose a way to study the function of the rods alone, given they can sense light about 1000 times dimmer than the cones.

Glossary

hues

identity of a color as it relates specifically to the spectrum

rods and cones

two types of photoreceptors in the human retina; rods are responsible for vision at low light levels, while cones are active at higher light levels

simplified theory of color vision

a theory that states that there are three primary colors, which correspond to the three types of cones

color constancy

a part of the visual perception system that allows people to perceive color in a variety of conditions and to see some consistency in the color

retinex

a theory proposed to explain color and brightness perception and constancies; is a combination of the words retina and cortex, which are the two areas responsible for the processing of visual information

retinex theory of color vision

the ability to perceive color in an ambient-colored environment

Microscopes

- Investigate different types of microscopes.
- Learn how image is formed in a compound microscope.

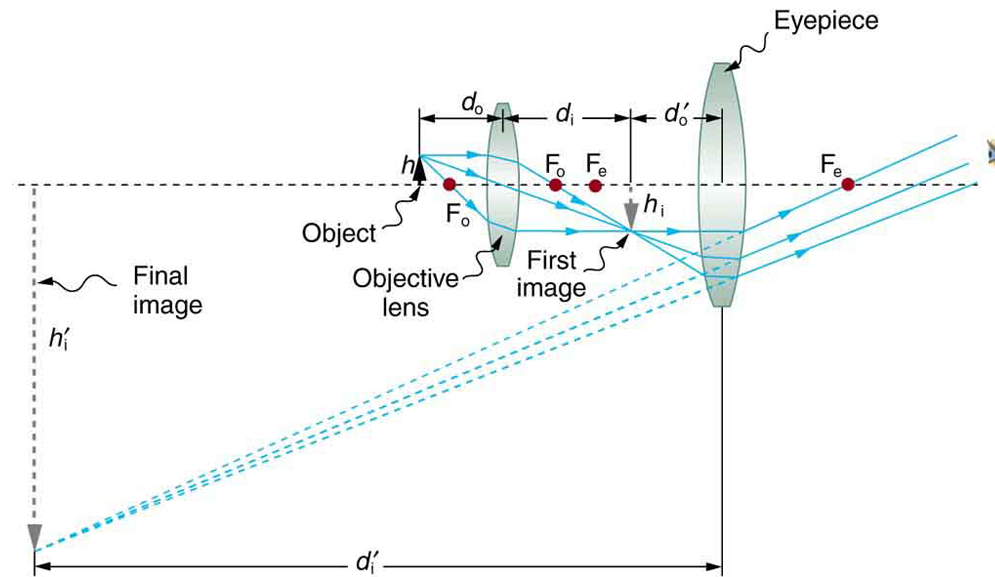
Although the eye is marvelous in its ability to see objects large and small, it obviously has limitations to the smallest details it can detect. Human desire to see beyond what is possible with the naked eye led to the use of optical instruments. In this section we will examine microscopes, instruments for enlarging the detail that we cannot see with the unaided eye. The microscope is a multiple-element system having more than a single lens or mirror. (See [\[link\]](#)) A microscope can be made from two convex lenses. The image formed by the first element becomes the object for the second element. The second element forms its own image, which is the object for the third element, and so on. Ray tracing helps to visualize the image formed. If the device is composed of thin lenses and mirrors that obey the thin lens equations, then it is not difficult to describe their behavior numerically.



Multiple lenses and mirrors are used in this microscope. (credit: U.S. Navy photo by Tom Watanabe)

Microscopes were first developed in the early 1600s by eyeglass makers in The Netherlands and Denmark. The simplest **compound microscope** is

constructed from two convex lenses as shown schematically in [\[link\]](#). The first lens is called the **objective lens**, and has typical magnification values from $5\times$ to $100\times$. In standard microscopes, the objectives are mounted such that when you switch between objectives, the sample remains in focus. Objectives arranged in this way are described as parfocal. The second, the **eyepiece**, also referred to as the ocular, has several lenses which slide inside a cylindrical barrel. The focusing ability is provided by the movement of both the objective lens and the eyepiece. The purpose of a microscope is to magnify small objects, and both lenses contribute to the final magnification. Additionally, the final enlarged image is produced in a location far enough from the observer to be easily viewed, since the eye cannot focus on objects or images that are too close.



A compound microscope composed of two lenses, an objective and an eyepiece. The objective forms a case 1 image that is larger than the object. This first image is the object for the eyepiece. The eyepiece forms a case 2 final image that is further magnified.

To see how the microscope in [\[link\]](#) forms an image, we consider its two lenses in succession. The object is slightly farther away from the objective lens than its focal length f_o , producing a case 1 image that is larger than the

object. This first image is the object for the second lens, or eyepiece. The eyepiece is intentionally located so it can further magnify the image. The eyepiece is placed so that the first image is closer to it than its focal length f_e . Thus the eyepiece acts as a magnifying glass, and the final image is made even larger. The final image remains inverted, but it is farther from the observer, making it easy to view (the eye is most relaxed when viewing distant objects and normally cannot focus closer than 25 cm). Since each lens produces a magnification that multiplies the height of the image, it is apparent that the overall magnification m is the product of the individual magnifications:

Equation:

$$m = m_o m_e,$$

where m_o is the magnification of the objective and m_e is the magnification of the eyepiece. This equation can be generalized for any combination of thin lenses and mirrors that obey the thin lens equations.

Note:

Overall Magnification

The overall magnification of a multiple-element system is the product of the individual magnifications of its elements.

Example:

Microscope Magnification

Calculate the magnification of an object placed 6.20 mm from a compound microscope that has a 6.00 mm focal length objective and a 50.0 mm focal length eyepiece. The objective and eyepiece are separated by 23.0 cm.

Strategy and Concept

This situation is similar to that shown in [\[link\]](#). To find the overall magnification, we must find the magnification of the objective, then the magnification of the eyepiece. This involves using the thin lens equation.

Solution

The magnification of the objective lens is given as

Equation:

$$m_o = -\frac{d_i}{d_o},$$

where d_o and d_i are the object and image distances, respectively, for the objective lens as labeled in [\[link\]](#). The object distance is given to be $d_o = 6.20$ mm, but the image distance d_i is not known. Isolating d_i , we have

Equation:

$$\frac{1}{d_i} = \frac{1}{f_o} - \frac{1}{d_o},$$

where f_o is the focal length of the objective lens. Substituting known values gives

Equation:

$$\frac{1}{d_i} = \frac{1}{6.00 \text{ mm}} - \frac{1}{6.20 \text{ mm}} = \frac{0.00538}{\text{mm}}.$$

We invert this to find d_i :

Equation:

$$d_i = 186 \text{ mm}.$$

Substituting this into the expression for m_o gives

Equation:

$$m_o = -\frac{d_i}{d_o} = -\frac{186 \text{ mm}}{6.20 \text{ mm}} = -30.0.$$

Now we must find the magnification of the eyepiece, which is given by

Equation:

$$m_e = -\frac{d_i'}{d_o'},$$

where d_i' and d_o' are the image and object distances for the eyepiece (see [\[link\]](#)). The object distance is the distance of the first image from the eyepiece. Since the first image is 186 mm to the right of the objective and the eyepiece is 230 mm to the right of the objective, the object distance is $d_o' = 230 \text{ mm} - 186 \text{ mm} = 44.0 \text{ mm}$. This places the first image closer to the eyepiece than its focal length, so that the eyepiece will form a case 2 image as shown in the figure. We still need to find the location of the final image d_i' in order to find the magnification. This is done as before to obtain a value for $1/d_i'$:

Equation:

$$\frac{1}{d_i'} = \frac{1}{f_e} - \frac{1}{d_o'} = \frac{1}{50.0 \text{ mm}} - \frac{1}{44.0 \text{ mm}} = -\frac{0.00273}{\text{mm}}.$$

Inverting gives

Equation:

$$d_i' = -\frac{\text{mm}}{0.00273} = -367 \text{ mm}.$$

The eyepiece's magnification is thus

Equation:

$$m_e = -\frac{d_i'}{d_o'} = -\frac{-367 \text{ mm}}{44.0 \text{ mm}} = 8.33.$$

So the overall magnification is

Equation:

$$m = m_o m_e = (-30.0)(8.33) = -250.$$

Discussion

Both the objective and the eyepiece contribute to the overall magnification, which is large and negative, consistent with [\[link\]](#), where the image is seen to be large and inverted. In this case, the image is virtual and inverted, which cannot happen for a single element (case 2 and case 3 images for single elements are virtual and upright). The final image is 367 mm (0.367 m) to the left of the eyepiece. Had the eyepiece been placed farther from

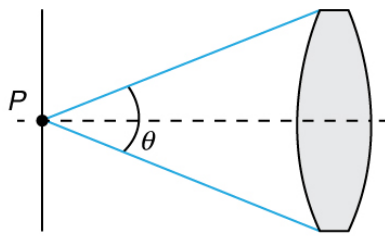
the objective, it could have formed a case 1 image to the right. Such an image could be projected on a screen, but it would be behind the head of the person in the figure and not appropriate for direct viewing. The procedure used to solve this example is applicable in any multiple-element system. Each element is treated in turn, with each forming an image that becomes the object for the next element. The process is not more difficult than for single lenses or mirrors, only lengthier.

Normal optical microscopes can magnify up to $1500\times$ with a theoretical resolution of $\sim 0.2\text{ }\mu\text{m}$. The lenses can be quite complicated and are composed of multiple elements to reduce aberrations. Microscope objective lenses are particularly important as they primarily gather light from the specimen. Three parameters describe microscope objectives: the **numerical aperture** (NA), the magnification (m), and the working distance. The NA is related to the light gathering ability of a lens and is obtained using the angle of acceptance θ formed by the maximum cone of rays focusing on the specimen (see [\[link\]](#)(a)) and is given by

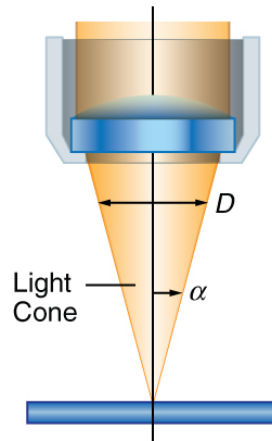
Equation:

$$\text{NA} = n \sin \alpha,$$

where n is the refractive index of the medium between the lens and the specimen and $\alpha = \theta/2$. As the angle of acceptance given by θ increases, NA becomes larger and more light is gathered from a smaller focal region giving higher resolution. A 0.75NA objective gives more detail than a 0.10NA objective.



(a)



(b)

(a) The numerical aperture (NA) of a microscope objective lens refers to the light-gathering ability of the lens and is calculated using half the angle of acceptance θ . (b) Here, α is half the acceptance angle for light rays from a specimen entering a camera lens, and D is the diameter of the aperture that controls the light entering the lens.

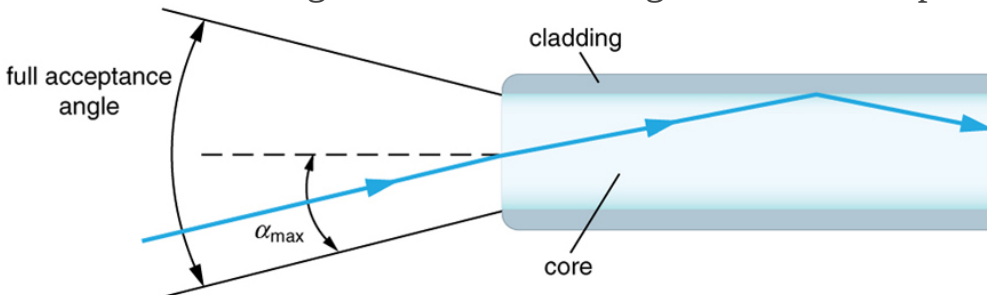
While the numerical aperture can be used to compare resolutions of various objectives, it does not indicate how far the lens could be from the specimen. This is specified by the “working distance,” which is the distance (in mm usually) from the front lens element of the objective to the specimen, or cover glass. The higher the NA the closer the lens will be to the specimen and the more chances there are of breaking the cover slip and damaging both the specimen and the lens. The focal length of an objective lens is different than the working distance. This is because objective lenses are made of a combination of lenses and the focal length is measured from inside the barrel. The working distance is a parameter that microscopists can use more readily as it is measured from the outermost lens. The working distance decreases as the NA and magnification both increase.

The term $f/\#$ in general is called the f -number and is used to denote the light per unit area reaching the image plane. In photography, an image of an object at infinity is formed at the focal point and the f -number is given by the ratio of the focal length f of the lens and the diameter D of the aperture controlling the light into the lens (see [\[link\]](#)(b)). If the acceptance angle is small the NA of the lens can also be used as given below.

Equation:

$$f/\# = \frac{f}{D} \approx \frac{1}{2\text{NA}}.$$

As the f -number decreases, the camera is able to gather light from a larger angle, giving wide-angle photography. As usual there is a trade-off. A greater $f/\#$ means less light reaches the image plane. A setting of $f/16$ usually allows one to take pictures in bright sunlight as the aperture diameter is small. In optical fibers, light needs to be focused into the fiber. [\[link\]](#) shows the angle used in calculating the NA of an optical fiber.

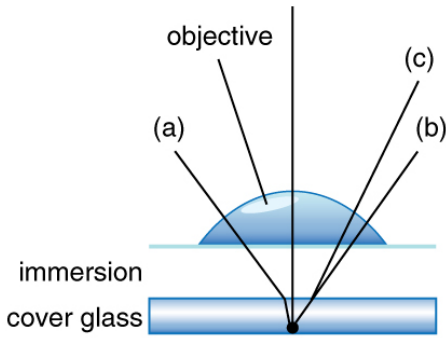


Light rays enter an optical fiber. The numerical aperture of the optical fiber can be determined by using the angle

$$\alpha_{\text{max}}.$$

Can the NA be larger than 1.00? The answer is ‘yes’ if we use immersion lenses in which a medium such as oil, glycerine or water is placed between the objective and the microscope cover slip. This minimizes the mismatch in refractive indices as light rays go through different media, generally providing a greater light-gathering ability and an increase in resolution.

[\[link\]](#) shows light rays when using air and immersion lenses.

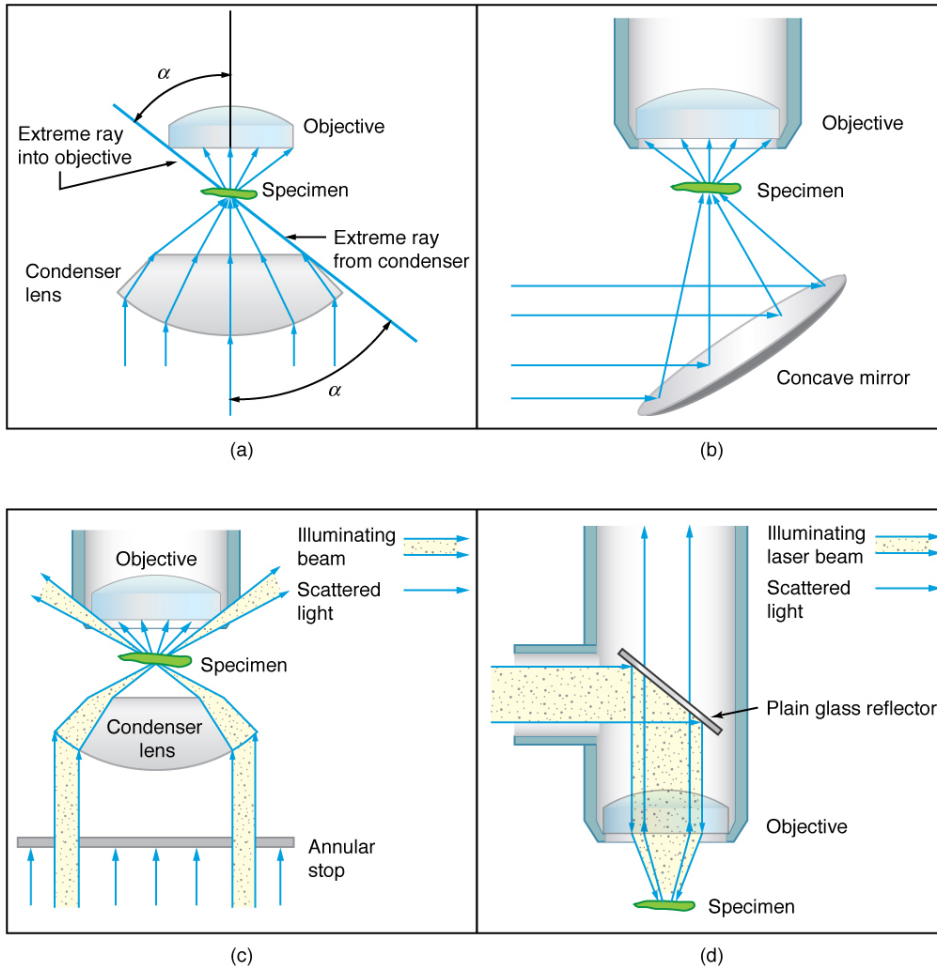


Light rays from a specimen entering the objective. Paths for immersion medium of air (a), water (b) ($n = 1.33$), and oil (c) ($n = 1.51$) are shown. The water and oil immersions allow more rays to enter the objective, increasing the resolution.

When using a microscope we do not see the entire extent of the sample. Depending on the eyepiece and objective lens we see a restricted region which we say is the field of view. The objective is then manipulated in two-dimensions above the sample to view other regions of the sample. Electronic scanning of either the objective or the sample is used in scanning microscopy. The image formed at each point during the scanning is combined using a computer to generate an image of a larger region of the sample at a selected magnification.

When using a microscope, we rely on gathering light to form an image. Hence most specimens need to be illuminated, particularly at higher magnifications, when observing details that are so small that they reflect only small amounts of light. To make such objects easily visible, the intensity of light falling on them needs to be increased. Special illuminating

systems called condensers are used for this purpose. The type of condenser that is suitable for an application depends on how the specimen is examined, whether by transmission, scattering or reflecting. See [\[link\]](#) for an example of each. White light sources are common and lasers are often used. Laser light illumination tends to be quite intense and it is important to ensure that the light does not result in the degradation of the specimen.

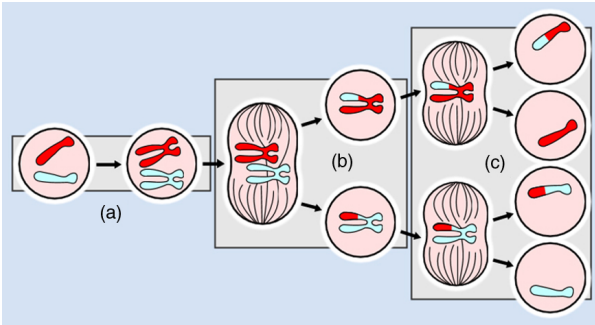


Illumination of a specimen in a microscope. (a)
 Transmitted light from a condenser lens. (b)
 Transmitted light from a mirror condenser. (c) Dark
 field illumination by scattering (the illuminating beam
 misses the objective lens). (d) High magnification
 illumination with reflected light – normally laser
 light.

We normally associate microscopes with visible light but x ray and electron microscopes provide greater resolution. The focusing and basic physics is the same as that just described, even though the lenses require different technology. The electron microscope requires vacuum chambers so that the electrons can proceed unheeded. Magnifications of 50 million times provide the ability to determine positions of individual atoms within materials. An electron microscope is shown in [\[link\]](#). We do not use our eyes to form images; rather images are recorded electronically and displayed on computers. In fact observing and saving images formed by optical microscopes on computers is now done routinely. Video recordings of what occurs in a microscope can be made for viewing by many people at later dates. Physics provides the science and tools needed to generate the sequence of time-lapse images of meiosis similar to the sequence sketched in [\[link\]](#).



An electron microscope
has the capability to
image individual atoms
on a material. The
microscope uses vacuum
technology, sophisticated
detectors and state of the
art image processing
software. (credit: Dave
Pape)



The image shows a sequence of events that takes place during meiosis. (credit: PatríciaR, Wikimedia Commons; National Center for Biotechnology Information)

Note:

Take-Home Experiment: Make a Lens

Look through a clear glass or plastic bottle and describe what you see. Now fill the bottle with water and describe what you see. Use the water bottle as a lens to produce the image of a bright object and estimate the focal length of the water bottle lens. How is the focal length a function of the depth of water in the bottle?

Section Summary

- The microscope is a multiple-element system having more than a single lens or mirror.
- Many optical devices contain more than a single lens or mirror. These are analysed by considering each element sequentially. The image formed by the first is the object for the second, and so on. The same ray tracing and thin lens techniques apply to each lens element.

- The overall magnification of a multiple-element system is the product of the magnifications of its individual elements. For a two-element system with an objective and an eyepiece, this is

Equation:

$$m = m_o m_e,$$

where m_o is the magnification of the objective and m_e is the magnification of the eyepiece, such as for a microscope.

- Microscopes are instruments for allowing us to see detail we would not be able to see with the unaided eye and consist of a range of components.
- The eyepiece and objective contribute to the magnification. The numerical aperture (NA) of an objective is given by

Equation:

$$NA = n \sin \alpha$$

where n is the refractive index and α the angle of acceptance.

- Immersion techniques are often used to improve the light gathering ability of microscopes. The specimen is illuminated by transmitted, scattered or reflected light through a condenser.
- The $f / \#$ describes the light gathering ability of a lens. It is given by

Equation:

$$f / \# = \frac{f}{D} \approx \frac{1}{2 NA}.$$

Conceptual Questions

Exercise:

Problem:

Geometric optics describes the interaction of light with macroscopic objects. Why, then, is it correct to use geometric optics to analyse a microscope's image?

Exercise:**Problem:**

The image produced by the microscope in [\[link\]](#) cannot be projected. Could extra lenses or mirrors project it? Explain.

Exercise:**Problem:**

Why not have the objective of a microscope form a case 2 image with a large magnification? (Hint: Consider the location of that image and the difficulty that would pose for using the eyepiece as a magnifier.)

Exercise:

Problem: What advantages do oil immersion objectives offer?

Exercise:**Problem:**

How does the NA of a microscope compare with the NA of an optical fiber?

Problem Exercises**Exercise:****Problem:**

A microscope with an overall magnification of 800 has an objective that magnifies by 200. (a) What is the magnification of the eyepiece? (b) If there are two other objectives that can be used, having magnifications of 100 and 400, what other total magnifications are possible?

Solution:

(a) 4.00

(b) 1600

Exercise:

Problem:

- (a) What magnification is produced by a 0.150 cm focal length microscope objective that is 0.155 cm from the object being viewed?
- (b) What is the overall magnification if an $8\times$ eyepiece (one that produces a magnification of 8.00) is used?

Exercise:

Problem:

- (a) Where does an object need to be placed relative to a microscope for its 0.500 cm focal length objective to produce a magnification of -400 ?
- (b) Where should the 5.00 cm focal length eyepiece be placed to produce a further fourfold (4.00) magnification?

Solution:

(a) 0.501 cm

(b) Eyepiece should be 204 cm behind the objective lens.

Exercise:

Problem:

You switch from a $1.40NA$ $60\times$ oil immersion objective to a $1.40NA$ $60\times$ oil immersion objective. What are the acceptance angles for each? Compare and comment on the values. Which would you use first to locate the target area on your specimen?

Exercise:

Problem:

An amoeba is 0.305 cm away from the 0.300 cm focal length objective lens of a microscope. (a) Where is the image formed by the objective lens? (b) What is this image's magnification? (c) An eyepiece with a 2.00 cm focal length is placed 20.0 cm from the objective. Where is the final image? (d) What magnification is produced by the eyepiece? (e) What is the overall magnification? (See [\[link\]](#).)

Solution:

- (a) +18.3 cm (on the eyepiece side of the objective lens)
- (b) -60.0
- (c) -11.3 cm (on the objective side of the eyepiece)
- (d) +6.67
- (e) -400

Exercise:**Problem:**

You are using a standard microscope with a $0.10\text{ NA } 4\times$ objective and switch to a $0.65\text{ NA } 40\times$ objective. What are the acceptance angles for each? Compare and comment on the values. Which would you use first to locate the target area on of your specimen? (See [\[link\]](#).)

Exercise:**Problem: Unreasonable Results**

Your friends show you an image through a microscope. They tell you that the microscope has an objective with a 0.500 cm focal length and an eyepiece with a 5.00 cm focal length. The resulting overall magnification is 250,000. Are these viable values for a microscope?

Glossary

compound microscope

a microscope constructed from two convex lenses, the first serving as the ocular lens(close to the eye) and the second serving as the objective lens

objective lens

the lens nearest to the object being examined

eyepiece

the lens or combination of lenses in an optical instrument nearest to the eye of the observer

numerical aperture

a number or measure that expresses the ability of a lens to resolve fine detail in an object being observed. Derived by mathematical formula

Equation:

$$NA = n \sin \alpha,$$

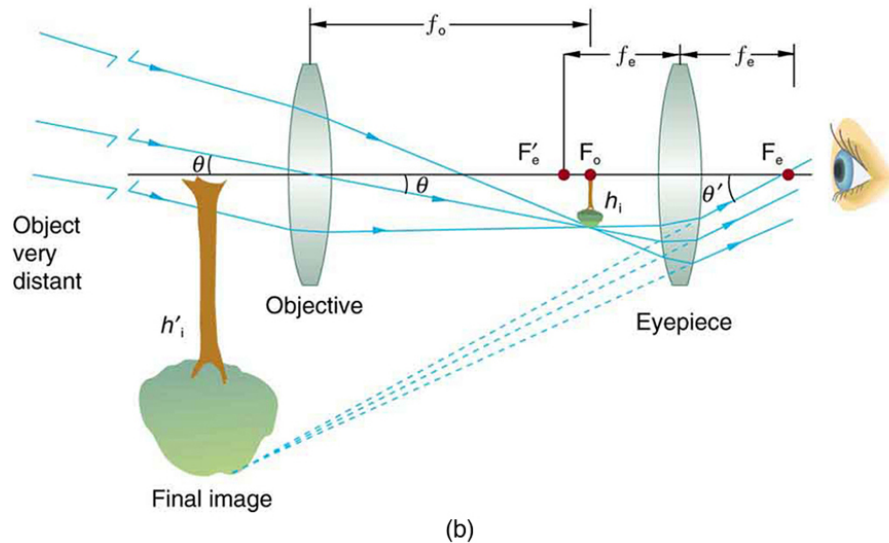
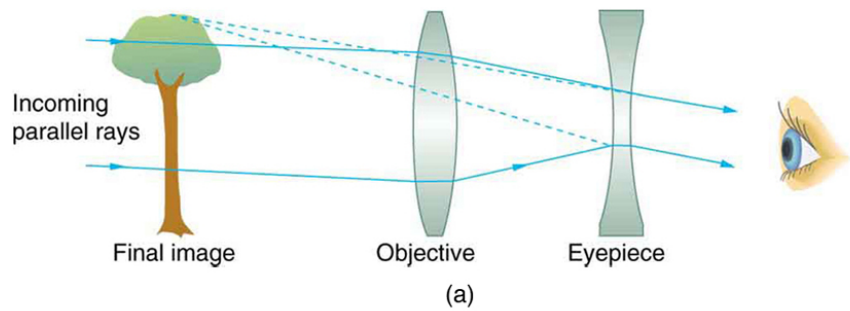
where n is the refractive index of the medium between the lens and the specimen and $\alpha = \theta/2$

Telescopes

- Outline the invention of a telescope.
- Describe the working of a telescope.

Telescopes are meant for viewing distant objects, producing an image that is larger than the image that can be seen with the unaided eye. Telescopes gather far more light than the eye, allowing dim objects to be observed with greater magnification and better resolution. Although Galileo is often credited with inventing the telescope, he actually did not. What he did was more important. He constructed several early telescopes, was the first to study the heavens with them, and made monumental discoveries using them. Among these are the moons of Jupiter, the craters and mountains on the Moon, the details of sunspots, and the fact that the Milky Way is composed of vast numbers of individual stars.

[\[link\]](#)(a) shows a telescope made of two lenses, the convex objective and the concave eyepiece, the same construction used by Galileo. Such an arrangement produces an upright image and is used in spyglasses and opera glasses.



(a) Galileo made telescopes with a convex objective and a concave eyepiece. These produce an upright image and are used in spyglasses. (b) Most simple telescopes have two convex lenses. The objective forms a case 1 image that is the object for the eyepiece. The eyepiece forms a case 2 final image that is magnified.

The most common two-lens telescope, like the simple microscope, uses two convex lenses and is shown in [\[link\]](#)(b). The object is so far away from the telescope that it is essentially at infinity compared with the focal lengths of the lenses ($d_o \approx \infty$). The first image is thus produced at $d_i = f_o$, as shown in the figure. To prove this, note that

Equation:

$$\frac{1}{d_i} = \frac{1}{f_o} - \frac{1}{d_o} = \frac{1}{f_o} - \frac{1}{\infty}.$$

Because $1/\infty = 0$, this simplifies to

Equation:

$$\frac{1}{d_i} = \frac{1}{f_o},$$

which implies that $d_i = f_o$, as claimed. It is true that for any distant object and any lens or mirror, the image is at the focal length.

The first image formed by a telescope objective as seen in [\[link\]](#)(b) will not be large compared with what you might see by looking at the object directly. For example, the spot formed by sunlight focused on a piece of paper by a magnifying glass is the image of the Sun, and it is small. The telescope eyepiece (like the microscope eyepiece) magnifies this first image. The distance between the eyepiece and the objective lens is made slightly less than the sum of their focal lengths so that the first image is closer to the eyepiece than its focal length. That is, d_o' is less than f_e , and so the eyepiece forms a case 2 image that is large and to the left for easy viewing. If the angle subtended by an object as viewed by the unaided eye is θ , and the angle subtended by the telescope image is θ' , then the **angular magnification** M is defined to be their ratio. That is, $M = \theta'/\theta$. It can be shown that the angular magnification of a telescope is related to the focal lengths of the objective and eyepiece; and is given by

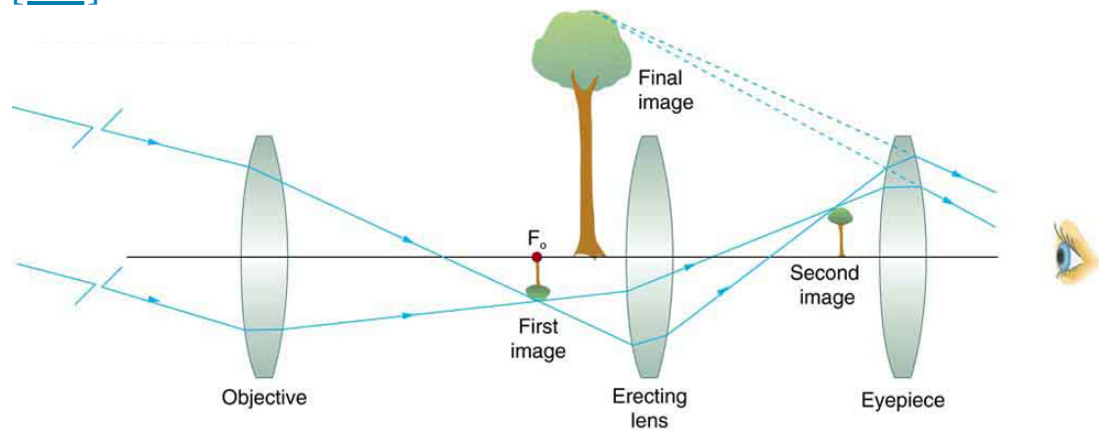
Equation:

$$M = \frac{\theta'}{\theta} = -\frac{f_o}{f_e}.$$

The minus sign indicates the image is inverted. To obtain the greatest angular magnification, it is best to have a long focal length objective and a short focal length eyepiece. The greater the angular magnification M , the larger an object will appear when viewed through a telescope, making more

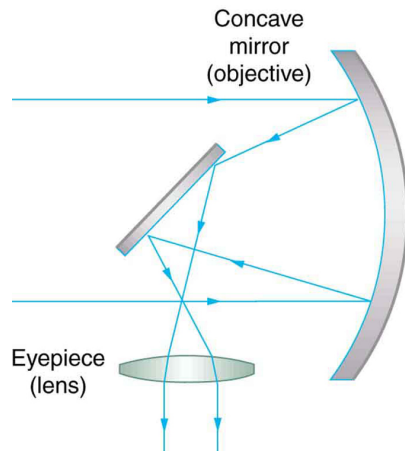
details visible. Limits to observable details are imposed by many factors, including lens quality and atmospheric disturbance.

The image in most telescopes is inverted, which is unimportant for observing the stars but a real problem for other applications, such as telescopes on ships or telescopic gun sights. If an upright image is needed, Galileo's arrangement in [\[link\]](#)(a) can be used. But a more common arrangement is to use a third convex lens as an eyepiece, increasing the distance between the first two and inverting the image once again as seen in [\[link\]](#).



This arrangement of three lenses in a telescope produces an upright final image. The first two lenses are far enough apart that the second lens inverts the image of the first one more time. The third lens acts as a magnifier and keeps the image upright and in a location that is easy to view.

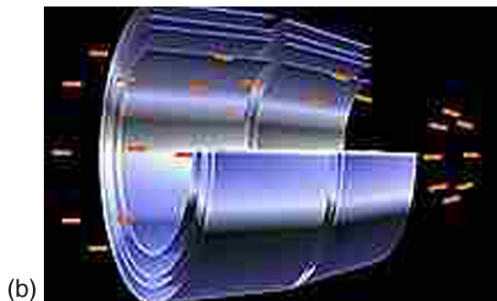
A telescope can also be made with a concave mirror as its first element or objective, since a concave mirror acts like a convex lens as seen in [\[link\]](#). Flat mirrors are often employed in optical instruments to make them more compact or to send light to cameras and other sensing devices. There are many advantages to using mirrors rather than lenses for telescope objectives. Mirrors can be constructed much larger than lenses and can, thus, gather large amounts of light, as needed to view distant galaxies, for example. Large and relatively flat mirrors have very long focal lengths, so that great angular magnification is possible.



A two-element telescope composed of a mirror as the objective and a lens for the eyepiece is shown. This telescope forms an image in the same manner as the two-convex-lens telescope already discussed, but it does not suffer from chromatic aberrations. Such telescopes can gather more light, since larger mirrors than lenses can be constructed.

Telescopes, like microscopes, can utilize a range of frequencies from the electromagnetic spectrum. [\[link\]](#)(a) shows the Australia Telescope Compact

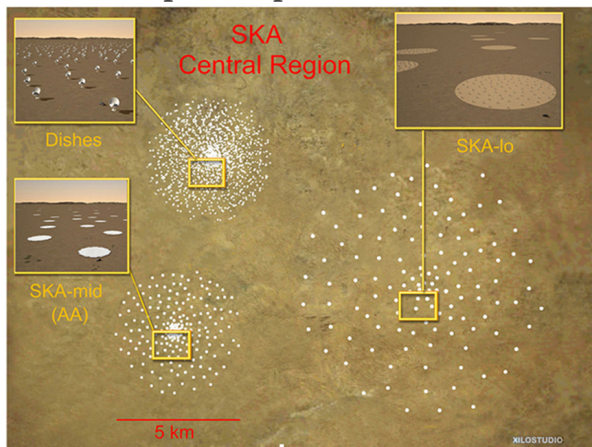
Array, which uses six 22-m antennas for mapping the southern skies using radio waves. [\[link\]](#)(b) shows the focusing of x rays on the Chandra X-ray Observatory—a satellite orbiting earth since 1999 and looking at high temperature events as exploding stars, quasars, and black holes. X rays, with much more energy and shorter wavelengths than RF and light, are mainly absorbed and not reflected when incident perpendicular to the medium. But they can be reflected when incident at small glancing angles, much like a rock will skip on a lake if thrown at a small angle. The mirrors for the Chandra consist of a long barrelled pathway and 4 pairs of mirrors to focus the rays at a point 10 meters away from the entrance. The mirrors are extremely smooth and consist of a glass ceramic base with a thin coating of metal (iridium). Four pairs of precision manufactured mirrors are exquisitely shaped and aligned so that x rays ricochet off the mirrors like bullets off a wall, focusing on a spot.



(a) The Australia Telescope Compact Array at Narrabri (500 km NW of Sydney). (credit: Ian

Bailey) (b) The focusing of x rays on the Chandra Observatory, a satellite orbiting earth. X rays ricochet off 4 pairs of mirrors forming a barrelled pathway leading to the focus point. (credit: NASA)

A current exciting development is a collaborative effort involving 17 countries to construct a Square Kilometre Array (SKA) of telescopes capable of covering from 80 MHz to 2 GHz. The initial stage of the project is the construction of the Australian Square Kilometre Array Pathfinder in Western Australia (see [\[link\]](#)). The project will use cutting-edge technologies such as **adaptive optics** in which the lens or mirror is constructed from lots of carefully aligned tiny lenses and mirrors that can be manipulated using computers. A range of rapidly changing distortions can be minimized by deforming or tilting the tiny lenses and mirrors. The use of adaptive optics in vision correction is a current area of research.



An artist's impression of the Australian Square Kilometre Array Pathfinder in Western

Australia is displayed. (credit: SPDO, XILOSTUDIOS)

Section Summary

- Simple telescopes can be made with two lenses. They are used for viewing objects at large distances and utilize the entire range of the electromagnetic spectrum.
- The angular magnification M for a telescope is given by
Equation:

$$M = \frac{\theta'}{\theta} = -\frac{f_o}{f_e},$$

where θ is the angle subtended by an object viewed by the unaided eye, θ' is the angle subtended by a magnified image, and f_o and f_e are the focal lengths of the objective and the eyepiece.

Conceptual Questions

Exercise:

Problem:

If you want your microscope or telescope to project a real image onto a screen, how would you change the placement of the eyepiece relative to the objective?

Problem Exercises

Unless otherwise stated, the lens-to-retina distance is 2.00 cm.

Exercise:

Problem:

What is the angular magnification of a telescope that has a 100 cm focal length objective and a 2.50 cm focal length eyepiece?

Solution:

−40.0

Exercise:**Problem:**

Find the distance between the objective and eyepiece lenses in the telescope in the above problem needed to produce a final image very far from the observer, where vision is most relaxed. Note that a telescope is normally used to view very distant objects.

Exercise:**Problem:**

A large reflecting telescope has an objective mirror with a 10.0 m radius of curvature. What angular magnification does it produce when a 3.00 m focal length eyepiece is used?

Solution:

−1.67

Exercise:**Problem:**

A small telescope has a concave mirror with a 2.00 m radius of curvature for its objective. Its eyepiece is a 4.00 cm focal length lens. (a) What is the telescope's angular magnification? (b) What angle is subtended by a 25,000 km diameter sunspot? (c) What is the angle of its telescopic image?

Exercise:

Problem:

A $7.5\times$ binocular produces an angular magnification of -7.50 , acting like a telescope. (Mirrors are used to make the image upright.) If the binoculars have objective lenses with a 75.0 cm focal length, what is the focal length of the eyepiece lenses?

Solution:

$+10.0\text{ cm}$

Exercise:**Problem: Construct Your Own Problem**

Consider a telescope of the type used by Galileo, having a convex objective and a concave eyepiece as illustrated in [\[link\]](#)(a). Construct a problem in which you calculate the location and size of the image produced. Among the things to be considered are the focal lengths of the lenses and their relative placements as well as the size and location of the object. Verify that the angular magnification is greater than one. That is, the angle subtended at the eye by the image is greater than the angle subtended by the object.

Glossary

adaptive optics

optical technology in which computers adjust the lenses and mirrors in a device to correct for image distortions

angular magnification

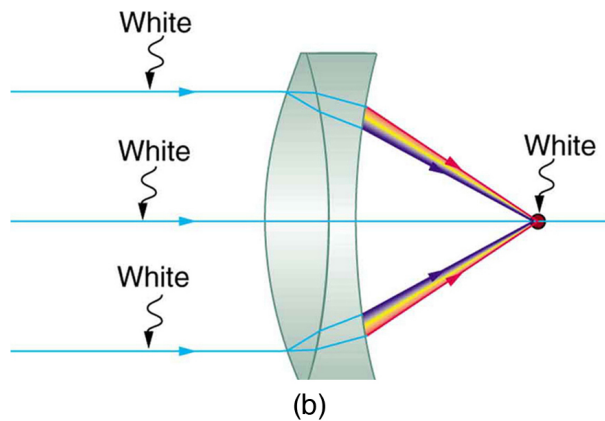
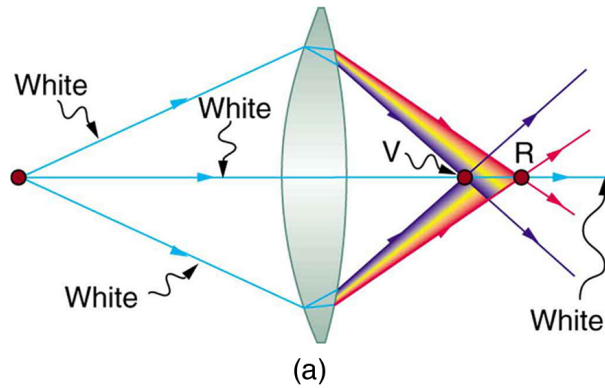
a ratio related to the focal lengths of the objective and eyepiece and given as $M = -\frac{f_o}{f_e}$

Aberrations

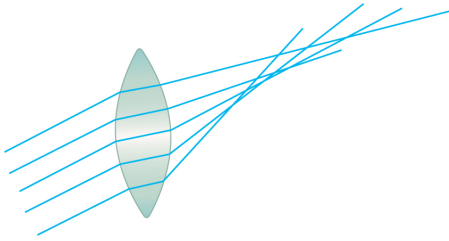
- Describe optical aberration.

Real lenses behave somewhat differently from how they are modeled using the thin lens equations, producing **aberrations**. An aberration is a distortion in an image. There are a variety of aberrations due to a lens size, material, thickness, and position of the object. One common type of aberration is chromatic aberration, which is related to color. Since the index of refraction of lenses depends on color or wavelength, images are produced at different places and with different magnifications for different colors. (The law of reflection is independent of wavelength, and so mirrors do not have this problem. This is another advantage for mirrors in optical systems such as telescopes.) [\[link\]](#)(a) shows chromatic aberration for a single convex lens and its partial correction with a two-lens system. Violet rays are bent more than red, since they have a higher index of refraction and are thus focused closer to the lens. The diverging lens partially corrects this, although it is usually not possible to do so completely. Lenses of different materials and having different dispersions may be used. For example an achromatic doublet consisting of a converging lens made of crown glass and a diverging lens made of flint glass in contact can dramatically reduce chromatic aberration (see [\[link\]](#)(b)).

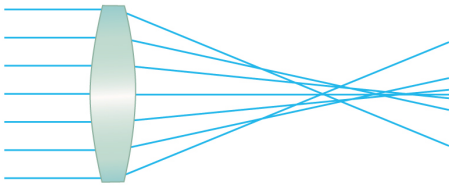
Quite often in an imaging system the object is off-center. Consequently, different parts of a lens or mirror do not refract or reflect the image to the same point. This type of aberration is called a coma and is shown in [\[link\]](#). The image in this case often appears pear-shaped. Another common aberration is spherical aberration where rays converging from the outer edges of a lens converge to a focus closer to the lens and rays closer to the axis focus further (see [\[link\]](#)). Aberrations due to astigmatism in the lenses of the eyes are discussed in [Vision Correction](#), and a chart used to detect astigmatism is shown in [\[link\]](#). Such aberrations and can also be an issue with manufactured lenses.



(a) Chromatic aberration is caused by the dependence of a lens's index of refraction on color (wavelength). The lens is more powerful for violet (V) than for red (R), producing images with different locations and magnifications. (b) Multiple-lens systems can partially correct chromatic aberrations, but they may require lenses of different materials and add to the expense of optical systems such as cameras.



A coma is an aberration caused by an object that is off-center, often resulting in a pear-shaped image. The rays originate from points that are not on the optical axis and they do not converge at one common focal point.



Spherical aberration is caused by rays focusing at different distances from the lens.

The image produced by an optical system needs to be bright enough to be discerned. It is often a challenge to obtain a sufficiently bright image. The brightness is determined by the amount of light passing through the optical system. The optical components determining the brightness are the diameter of the lens and the diameter of pupils, diaphragms or aperture stops placed

in front of lenses. Optical systems often have entrance and exit pupils to specifically reduce aberrations but they inevitably reduce brightness as well. Consequently, optical systems need to strike a balance between the various components used. The iris in the eye dilates and constricts, acting as an entrance pupil. You can see objects more clearly by looking through a small hole made with your hand in the shape of a fist. Squinting, or using a small hole in a piece of paper, also will make the object sharper.

So how are aberrations corrected? The lenses may also have specially shaped surfaces, as opposed to the simple spherical shape that is relatively easy to produce. Expensive camera lenses are large in diameter, so that they can gather more light, and need several elements to correct for various aberrations. Further, advances in materials science have resulted in lenses with a range of refractive indices—technically referred to as graded index (GRIN) lenses. Spectacles often have the ability to provide a range of focusing ability using similar techniques. GRIN lenses are particularly important at the end of optical fibers in endoscopes. Advanced computing techniques allow for a range of corrections on images after the image has been collected and certain characteristics of the optical system are known. Some of these techniques are sophisticated versions of what are available on commercial packages like Adobe Photoshop.

Section Summary

- Aberrations or image distortions can arise due to the finite thickness of optical instruments, imperfections in the optical components, and limitations on the ways in which the components are used.
- The means for correcting aberrations range from better components to computational techniques.

Conceptual Questions

Exercise:

Problem:

List the various types of aberrations. What causes them and how can each be reduced?

Problem Exercises**Exercise:****Problem: Integrated Concepts**

(a) During laser vision correction, a brief burst of 193 nm ultraviolet light is projected onto the cornea of the patient. It makes a spot 1.00 mm in diameter and deposits 0.500 mJ of energy. Calculate the depth of the layer ablated, assuming the corneal tissue has the same properties as water and is initially at 34.0°C. The tissue's temperature is increased to 100°C and evaporated without further temperature increase.

(b) Does your answer imply that the shape of the cornea can be finely controlled?

Solution:

(a) 0.251 μm

(b) Yes, this thickness implies that the shape of the cornea can be very finely controlled, producing normal distant vision in more than 90% of patients.

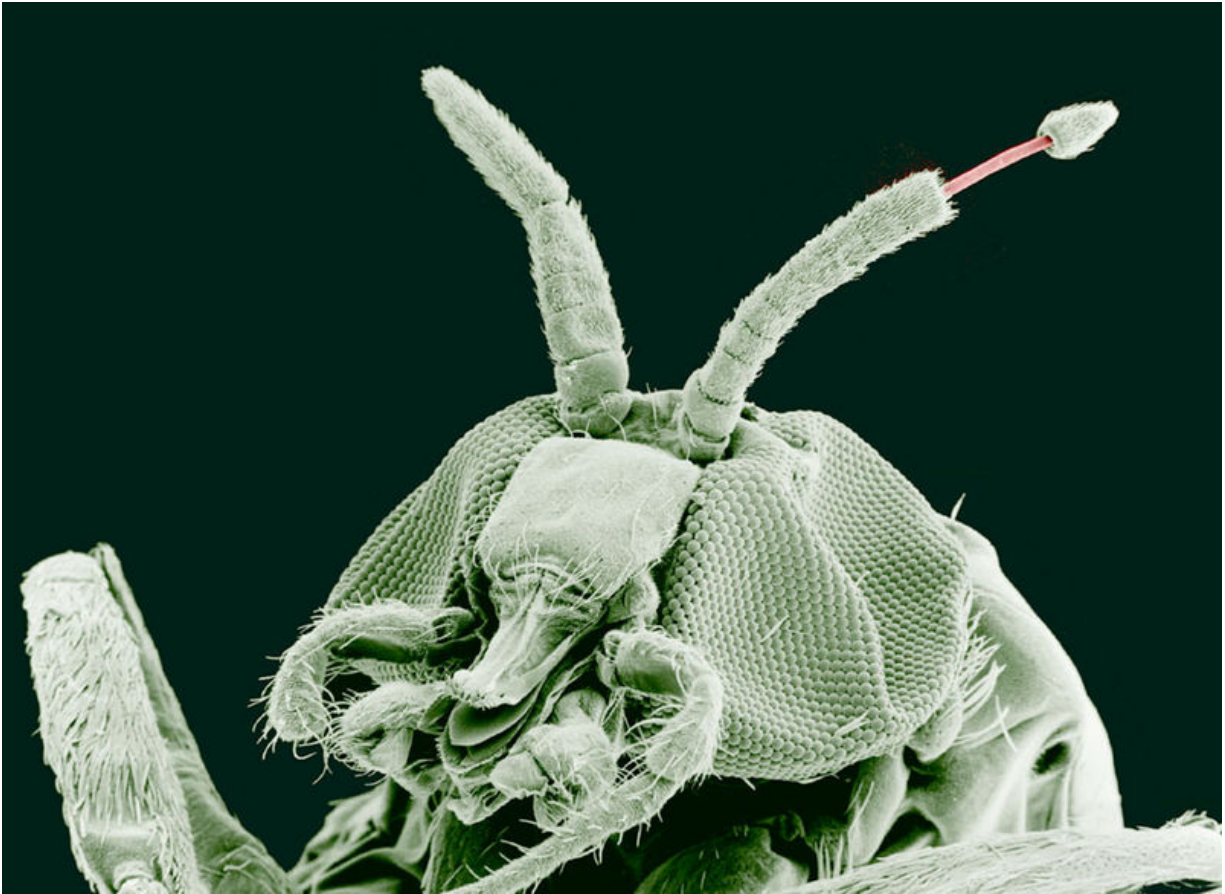
Glossary**aberration**

failure of rays to converge at one focus because of limitations or defects in a lens or mirror

Introduction to Quantum Physics

class="introduction"

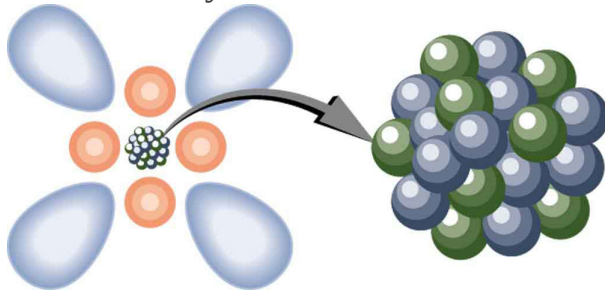
A black fly
imaged by
an electron
microscope
is as
monstrous
as any
science-
fiction
creature.
(credit:
U.S.
Departmen
t of
Agriculture
via
Wikimedia
Commons)



Quantum mechanics is the branch of physics needed to deal with submicroscopic objects. Because these objects are smaller than we can observe directly with our senses and generally must be observed with the aid of instruments, parts of quantum mechanics seem as foreign and bizarre as parts of relativity. But, like relativity, quantum mechanics has been shown to be valid—truth is often stranger than fiction.

Certain aspects of quantum mechanics are familiar to us. We accept as fact that matter is composed of atoms, the smallest unit of an element, and that these atoms combine to form molecules, the smallest unit of a compound. (See [\[link\]](#).) While we cannot see the individual water molecules in a stream, for example, we are aware that this is because molecules are so small and so numerous in that stream. When introducing atoms, we commonly say that electrons orbit atoms in discrete shells around a tiny nucleus, itself composed of smaller particles called protons and neutrons. We are also aware that electric charge comes in tiny units carried almost entirely by electrons and protons. As with water molecules in a stream, we

do not notice individual charges in the current through a lightbulb, because the charges are so small and so numerous in the macroscopic situations we sense directly.



Atoms and their substructure are familiar examples of objects that require quantum mechanics to be fully explained. Certain of their characteristics, such as the discrete electron shells, are classical physics explanations.

In quantum mechanics we conceptualize discrete “electron clouds” around the nucleus.

Note:

Making Connections: Realms of Physics

Classical physics is a good approximation of modern physics under conditions first discussed in the [The Nature of Science and Physics](#).

Quantum mechanics is valid in general, and it must be used rather than classical physics to describe small objects, such as atoms.

Atoms, molecules, and fundamental electron and proton charges are all examples of physical entities that are **quantized**—that is, they appear only in certain discrete values and do not have every conceivable value.

Quantized is the opposite of continuous. We cannot have a fraction of an atom, or part of an electron's charge, or 14-1/3 cents, for example. Rather, everything is built of integral multiples of these substructures. Quantum physics is the branch of physics that deals with small objects and the quantization of various entities, including energy and angular momentum. Just as with classical physics, quantum physics has several subfields, such as mechanics and the study of electromagnetic forces. The **correspondence principle** states that in the classical limit (large, slow-moving objects), **quantum mechanics** becomes the same as classical physics. In this chapter, we begin the development of quantum mechanics and its description of the strange submicroscopic world. In later chapters, we will examine many areas, such as atomic and nuclear physics, in which quantum mechanics is crucial.

Glossary

quantized

the fact that certain physical entities exist only with particular discrete values and not every conceivable value

correspondence principle

in the classical limit (large, slow-moving objects), quantum mechanics becomes the same as classical physics

quantum mechanics

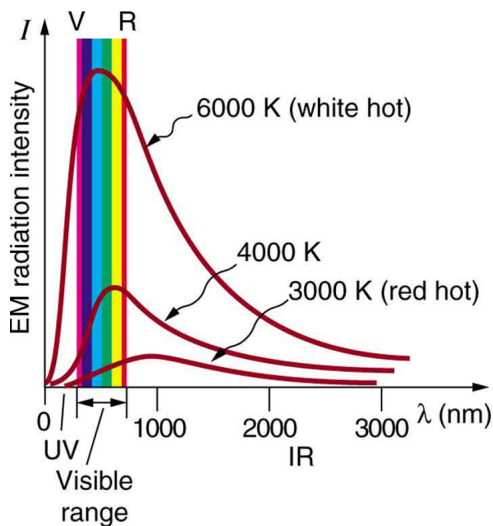
the branch of physics that deals with small objects and with the quantization of various entities, especially energy

Quantization of Energy

- Explain Max Planck's contribution to the development of quantum mechanics.
- Explain why atomic spectra indicate quantization.

Planck's Contribution

Energy is quantized in some systems, meaning that the system can have only certain energies and not a continuum of energies, unlike the classical case. This would be like having only certain speeds at which a car can travel because its kinetic energy can have only certain values. We also find that some forms of energy transfer take place with discrete lumps of energy. While most of us are familiar with the quantization of matter into lumps called atoms, molecules, and the like, we are less aware that energy, too, can be quantized. Some of the earliest clues about the necessity of quantum mechanics over classical physics came from the quantization of energy.



Graphs of blackbody radiation (from an ideal radiator) at three different radiator temperatures. The intensity or rate of

radiation emission increases dramatically with temperature, and the peak of the spectrum shifts toward the visible and ultraviolet parts of the spectrum. The shape of the spectrum cannot be described with classical physics.

Where is the quantization of energy observed? Let us begin by considering the emission and absorption of electromagnetic (EM) radiation. The EM spectrum radiated by a hot solid is linked directly to the solid's temperature. (See [\[link\]](#).) An ideal radiator is one that has an emissivity of 1 at all wavelengths and, thus, is jet black. Ideal radiators are therefore called **blackbodies**, and their EM radiation is called **blackbody radiation**. It was discussed that the total intensity of the radiation varies as T^4 , the fourth power of the absolute temperature of the body, and that the peak of the spectrum shifts to shorter wavelengths at higher temperatures. All of this seems quite continuous, but it was the curve of the spectrum of intensity versus wavelength that gave a clue that the energies of the atoms in the solid are quantized. In fact, providing a theoretical explanation for the experimentally measured shape of the spectrum was a mystery at the turn of the century. When this “ultraviolet catastrophe” was eventually solved, the answers led to new technologies such as computers and the sophisticated imaging techniques described in earlier chapters. Once again, physics as an enabling science changed the way we live.

The German physicist Max Planck (1858–1947) used the idea that atoms and molecules in a body act like oscillators to absorb and emit radiation. The energies of the oscillating atoms and molecules had to be quantized to correctly describe the shape of the blackbody spectrum. Planck deduced that the energy of an oscillator having a frequency f is given by

Equation:

$$E = \left(n + \frac{1}{2} \right) hf.$$

Here n is any nonnegative integer (0, 1, 2, 3, ...). The symbol h stands for **Planck's constant**, given by

Equation:

$$h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}.$$

The equation $E = \left(n + \frac{1}{2} \right) hf$ means that an oscillator having a frequency f (emitting and absorbing EM radiation of frequency f) can have its energy increase or decrease only in *discrete* steps of size

Equation:

$$\Delta E = hf.$$

It might be helpful to mention some macroscopic analogies of this quantization of energy phenomena. This is like a pendulum that has a characteristic oscillation frequency but can swing with only certain amplitudes. Quantization of energy also resembles a standing wave on a string that allows only particular harmonics described by integers. It is also similar to going up and down a hill using discrete stair steps rather than being able to move up and down a continuous slope. Your potential energy takes on discrete values as you move from step to step.

Using the quantization of oscillators, Planck was able to correctly describe the experimentally known shape of the blackbody spectrum. This was the first indication that energy is sometimes quantized on a small scale and earned him the Nobel Prize in Physics in 1918. Although Planck's theory comes from observations of a macroscopic object, its analysis is based on atoms and molecules. It was such a revolutionary departure from classical physics that Planck himself was reluctant to accept his own idea that energy states are not continuous. The general acceptance of Planck's energy quantization was greatly enhanced by Einstein's explanation of the photoelectric effect (discussed in the next section), which took energy

quantization a step further. Planck was fully involved in the development of both early quantum mechanics and relativity. He quickly embraced Einstein's special relativity, published in 1905, and in 1906 Planck was the first to suggest the correct formula for relativistic momentum, $p = \gamma mu$.



The German physicist Max Planck had a major influence on the early development of quantum mechanics, being the first to recognize that energy is sometimes quantized. Planck also made important contributions to special relativity and classical physics.
(credit: Library of Congress, Prints and Photographs Division via Wikimedia Commons)

Note that Planck's constant h is a very small number. So for an infrared frequency of 10^{14} Hz being emitted by a blackbody, for example, the difference between energy levels is only $\Delta E = hf = (6.63 \times 10^{-34} \text{ J}\cdot\text{s})(10^{14} \text{ Hz}) = 6.63 \times 10^{-20} \text{ J}$, or about 0.4 eV. This 0.4 eV of energy is significant compared with typical atomic

energies, which are on the order of an electron volt, or thermal energies, which are typically fractions of an electron volt. But on a macroscopic or classical scale, energies are typically on the order of joules. Even if macroscopic energies are quantized, the quantum steps are too small to be noticed. This is an example of the correspondence principle. For a large object, quantum mechanics produces results indistinguishable from those of classical physics.

Atomic Spectra

Now let us turn our attention to the *emission and absorption of EM radiation by gases*. The Sun is the most common example of a body containing gases emitting an EM spectrum that includes visible light. We also see examples in neon signs and candle flames. Studies of emissions of hot gases began more than two centuries ago, and it was soon recognized that these emission spectra contained huge amounts of information. The type of gas and its temperature, for example, could be determined. We now know that these EM emissions come from electrons transitioning between energy levels in individual atoms and molecules; thus, they are called **atomic spectra**. Atomic spectra remain an important analytical tool today. [\[link\]](#) shows an example of an emission spectrum obtained by passing an electric discharge through a material. One of the most important characteristics of these spectra is that they are discrete. By this we mean that only certain wavelengths, and hence frequencies, are emitted. This is called a line spectrum. If frequency and energy are associated as $\Delta E = hf$, the energies of the electrons in the emitting atoms and molecules are quantized. This is discussed in more detail later in this chapter.



Emission spectrum of oxygen. When an electrical discharge is passed through a substance, its atoms and molecules absorb energy, which is reemitted as EM radiation. The discrete nature of these emissions implies that the energy states of the atoms

and molecules are quantized. Such atomic spectra were used as analytical tools for many decades before it was understood why they are quantized. (credit: Teravolt, Wikimedia Commons)

It was a major puzzle that atomic spectra are quantized. Some of the best minds of 19th-century science failed to explain why this might be. Not until the second decade of the 20th century did an answer based on quantum mechanics begin to emerge. Again a macroscopic or classical body of gas was involved in the studies, but the effect, as we shall see, is due to individual atoms and molecules.

Note:

PhET Explorations: Models of the Hydrogen Atom

How did scientists figure out the structure of atoms without looking at them? Try out different models by shooting light at the atom. Check how the prediction of the model matches the experimental results.

<https://archive.cnx.org/specials/d77cc1d0-33e4-11e6-b016-6726afecd2be/hydrogen-atom/#sim-hydrogen-atom>

Section Summary

- The first indication that energy is sometimes quantized came from blackbody radiation, which is the emission of EM radiation by an object with an emissivity of 1.
- Planck recognized that the energy levels of the emitting atoms and molecules were quantized, with only the allowed values of $E = (n + \frac{1}{2})hf$, where n is any non-negative integer (0, 1, 2, 3, ...).
- h is Planck's constant, whose value is $h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$.
- Thus, the oscillatory absorption and emission energies of atoms and molecules in a blackbody could increase or decrease only in steps of

size $\Delta E = hf$ where f is the frequency of the oscillatory nature of the absorption and emission of EM radiation.

- Another indication of energy levels being quantized in atoms and molecules comes from the lines in atomic spectra, which are the EM emissions of individual atoms and molecules.

Conceptual Questions

Exercise:

Problem:

Give an example of a physical entity that is quantized. State specifically what the entity is and what the limits are on its values.

Exercise:

Problem:

Give an example of a physical entity that is not quantized, in that it is continuous and may have a continuous range of values.

Exercise:

Problem:

What aspect of the blackbody spectrum forced Planck to propose quantization of energy levels in its atoms and molecules?

Exercise:

Problem:

If Planck's constant were large, say 10^{34} times greater than it is, we would observe macroscopic entities to be quantized. Describe the motions of a child's swing under such circumstances.

Exercise:

Problem: Why don't we notice quantization in everyday events?

Problems & Exercises

Exercise:

Problem:

A LiBr molecule oscillates with a frequency of 1.7×10^{13} Hz. (a) What is the difference in energy in eV between allowed oscillator states? (b) What is the approximate value of n for a state having an energy of 1.0 eV?

Solution:

(a) 0.070 eV

(b) 14

Exercise:

Problem:

The difference in energy between allowed oscillator states in HBr molecules is 0.330 eV. What is the oscillation frequency of this molecule?

Exercise:

Problem:

A physicist is watching a 15-kg orangutan at a zoo swing lazily in a tire at the end of a rope. He (the physicist) notices that each oscillation takes 3.00 s and hypothesizes that the energy is quantized. (a) What is the difference in energy in joules between allowed oscillator states? (b) What is the value of n for a state where the energy is 5.00 J? (c) Can the quantization be observed?

Solution:

(a) 2.21×10^{-34} J

(b) 2.26×10^{34}

(c) No

Glossary

blackbody

an ideal radiator, which can radiate equally well at all wavelengths

blackbody radiation

the electromagnetic radiation from a blackbody

Planck's constant

$$h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$$

atomic spectra

the electromagnetic emission from atoms and molecules

The Photoelectric Effect

- Describe a typical photoelectric-effect experiment.
- Determine the maximum kinetic energy of photoelectrons ejected by photons of one energy or wavelength, when given the maximum kinetic energy of photoelectrons for a different photon energy or wavelength.

When light strikes materials, it can eject electrons from them. This is called the **photoelectric effect**, meaning that light (*photo*) produces electricity. One common use of the photoelectric effect is in light meters, such as those that adjust the automatic iris on various types of cameras. In a similar way, another use is in solar cells, as you probably have in your calculator or have seen on a roof top or a roadside sign. These make use of the photoelectric effect to convert light into electricity for running different devices.



The
photoelectric
effect can be
observed by
allowing
light to fall
on the metal
plate in this
evacuated
tube.

Electrons
ejected by
the light are
collected on
the collector
wire and

measured as
a current. A
retarding
voltage
between the
collector
wire and
plate can
then be
adjusted so
as to
determine the
energy of the
ejected
electrons. For
example, if it
is sufficiently
negative, no
electrons will
reach the
wire. (credit:
P.P. Urone)

This effect has been known for more than a century and can be studied using a device such as that shown in [\[link\]](#). This figure shows an evacuated tube with a metal plate and a collector wire that are connected by a variable voltage source, with the collector more negative than the plate. When light (or other EM radiation) strikes the plate in the evacuated tube, it may eject electrons. If the electrons have energy in electron volts (eV) greater than the potential difference between the plate and the wire in volts, some electrons will be collected on the wire. Since the electron energy in eV is qV , where q is the electron charge and V is the potential difference, the electron energy can be measured by adjusting the retarding voltage between the wire and the plate. The voltage that stops the electrons from reaching the wire equals the energy in eV. For example, if -3.00 V barely stops the electrons,

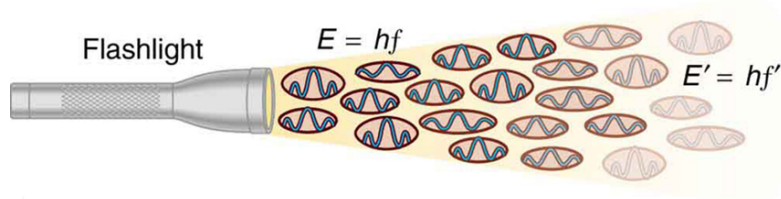
their energy is 3.00 eV. The number of electrons ejected can be determined by measuring the current between the wire and plate. The more light, the more electrons; a little circuitry allows this device to be used as a light meter.

What is really important about the photoelectric effect is what Albert Einstein deduced from it. Einstein realized that there were several characteristics of the photoelectric effect that could be explained only if *EM radiation is itself quantized*: the apparently continuous stream of energy in an EM wave is actually composed of energy quanta called photons. In his explanation of the photoelectric effect, Einstein defined a quantized unit or quantum of EM energy, which we now call a **photon**, with an energy proportional to the frequency of EM radiation. In equation form, the **photon energy** is

Equation:

$$E = hf,$$

where E is the energy of a photon of frequency f and h is Planck's constant. This revolutionary idea looks similar to Planck's quantization of energy states in blackbody oscillators, but it is quite different. It is the quantization of EM radiation itself. EM waves are composed of photons and are not continuous smooth waves as described in previous chapters on optics. Their energy is absorbed and emitted in lumps, not continuously. This is exactly consistent with Planck's quantization of energy levels in blackbody oscillators, since these oscillators increase and decrease their energy in steps of hf by absorbing and emitting photons having $E = hf$. We do not observe this with our eyes, because there are so many photons in common light sources that individual photons go unnoticed. (See [\[link\]](#).) The next section of the text ([Photon Energies and the Electromagnetic Spectrum](#)) is devoted to a discussion of photons and some of their characteristics and implications. For now, we will use the photon concept to explain the photoelectric effect, much as Einstein did.



An EM wave of frequency f is composed of photons, or individual quanta of EM radiation. The energy of each photon is $E = hf$, where h is Planck's constant and f is the frequency of the EM radiation. Higher intensity means more photons per unit area.

The flashlight emits large numbers of photons of many different frequencies, hence others have energy $E' = hf'$, and so on.

The photoelectric effect has the properties discussed below. All these properties are consistent with the idea that individual photons of EM radiation are absorbed by individual electrons in a material, with the electron gaining the photon's energy. Some of these properties are inconsistent with the idea that EM radiation is a simple wave. For simplicity, let us consider what happens with monochromatic EM radiation in which all photons have the same energy hf .

1. If we vary the frequency of the EM radiation falling on a material, we find the following: For a given material, there is a threshold frequency f_0 for the EM radiation below which no electrons are ejected, regardless of intensity. Individual photons interact with individual electrons. Thus if the photon energy is too small to break an electron away, no electrons will be ejected. If EM radiation was a simple wave, sufficient energy could be obtained by increasing the intensity.
2. *Once EM radiation falls on a material, electrons are ejected without delay.* As soon as an individual photon of a sufficiently high frequency is absorbed by an individual electron, the electron is ejected. If the EM

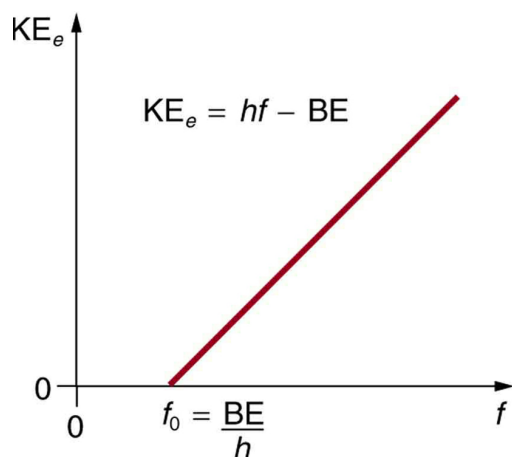
radiation were a simple wave, several minutes would be required for sufficient energy to be deposited to the metal surface to eject an electron.

3. The number of electrons ejected per unit time is proportional to the intensity of the EM radiation and to no other characteristic. High-intensity EM radiation consists of large numbers of photons per unit area, with all photons having the same characteristic energy hf .
4. If we vary the intensity of the EM radiation and measure the energy of ejected electrons, we find the following: *The maximum kinetic energy of ejected electrons is independent of the intensity of the EM radiation.* Since there are so many electrons in a material, it is extremely unlikely that two photons will interact with the same electron at the same time, thereby increasing the energy given it. Instead (as noted in 3 above), increased intensity results in more electrons of the same energy being ejected. If EM radiation were a simple wave, a higher intensity could give more energy, and higher-energy electrons would be ejected.
5. The kinetic energy of an ejected electron equals the photon energy minus the binding energy of the electron in the specific material. An individual photon can give all of its energy to an electron. The photon's energy is partly used to break the electron away from the material. The remainder goes into the ejected electron's kinetic energy. In equation form, this is given by

Equation:

$$KE_e = hf - BE,$$

where KE_e is the maximum kinetic energy of the ejected electron, hf is the photon's energy, and BE is the **binding energy** of the electron to the particular material. (BE is sometimes called the *work function* of the material.) This equation, due to Einstein in 1905, explains the properties of the photoelectric effect quantitatively. An individual photon of EM radiation (it does not come any other way) interacts with an individual electron, supplying enough energy, BE , to break it away, with the remainder going to kinetic energy. The binding energy is $BE = hf_0$, where f_0 is the threshold frequency for the particular material. [\[link\]](#) shows a graph of maximum KE_e versus the frequency of incident EM radiation falling on a particular material.



Photoelectric effect. A graph of the kinetic energy of an ejected electron, KE_e , versus the frequency of EM radiation impinging on a certain material. There is a threshold frequency below which no electrons are ejected, because the individual photon interacting with an individual electron has insufficient energy to break it away. Above the threshold energy, KE_e increases linearly with f , consistent with $KE_e = hf - BE$. The slope of this line is h — the data can be used to determine Planck's constant experimentally. Einstein gave the first successful explanation of such data by proposing

the idea of photons—
quanta of EM radiation.

Einstein's idea that EM radiation is quantized was crucial to the beginnings of quantum mechanics. It is a far more general concept than its explanation of the photoelectric effect might imply. All EM radiation can also be modeled in the form of photons, and the characteristics of EM radiation are entirely consistent with this fact. (As we will see in the next section, many aspects of EM radiation, such as the hazards of ultraviolet (UV) radiation, can be explained *only* by photon properties.) More famous for modern relativity, Einstein planted an important seed for quantum mechanics in 1905, the same year he published his first paper on special relativity. His explanation of the photoelectric effect was the basis for the Nobel Prize awarded to him in 1921. Although his other contributions to theoretical physics were also noted in that award, special and general relativity were not fully recognized in spite of having been partially verified by experiment by 1921. Although hero-worshipped, this great man never received Nobel recognition for his most famous work—relativity.

Example:

Calculating Photon Energy and the Photoelectric Effect: A Violet Light

(a) What is the energy in joules and electron volts of a photon of 420-nm violet light? (b) What is the maximum kinetic energy of electrons ejected from calcium by 420-nm violet light, given that the binding energy (or work function) of electrons for calcium metal is 2.71 eV?

Strategy

To solve part (a), note that the energy of a photon is given by $E = hf$. For part (b), once the energy of the photon is calculated, it is a straightforward application of $KE_e = hf - BE$ to find the ejected electron's maximum kinetic energy, since BE is given.

Solution for (a)

Photon energy is given by

Equation:

$$E = hf$$

Since we are given the wavelength rather than the frequency, we solve the familiar relationship $c = f\lambda$ for the frequency, yielding

Equation:

$$f = \frac{c}{\lambda}.$$

Combining these two equations gives the useful relationship

Equation:

$$E = \frac{hc}{\lambda}.$$

Now substituting known values yields

Equation:

$$E = \frac{(6.63 \times 10^{-34} \text{ J} \cdot \text{s})(3.00 \times 10^8 \text{ m/s})}{420 \times 10^{-9} \text{ m}} = 4.74 \times 10^{-19} \text{ J}.$$

Converting to eV, the energy of the photon is

Equation:

$$E = (4.74 \times 10^{-19} \text{ J}) \frac{1 \text{ eV}}{1.6 \times 10^{-19} \text{ J}} = 2.96 \text{ eV}.$$

Solution for (b)

Finding the kinetic energy of the ejected electron is now a simple application of the equation $\text{KE}_e = hf - \text{BE}$. Substituting the photon energy and binding energy yields

Equation:

$$\text{KE}_e = hf - \text{BE} = 2.96 \text{ eV} - 2.71 \text{ eV} = 0.246 \text{ eV}.$$

Discussion

The energy of this 420-nm photon of violet light is a tiny fraction of a joule, and so it is no wonder that a single photon would be difficult for us to sense directly—humans are more attuned to energies on the order of joules. But looking at the energy in electron volts, we can see that this photon has enough energy to affect atoms and molecules. A DNA molecule can be broken with about 1 eV of energy, for example, and typical atomic and molecular energies are on the order of eV, so that the UV photon in this example could have biological effects. The ejected electron (called a *photoelectron*) has a rather low energy, and it would not travel far, except in a vacuum. The electron would be stopped by a retarding potential of but 0.26 eV. In fact, if the photon wavelength were longer and its energy less than 2.71 eV, then the formula would give a negative kinetic energy, an impossibility. This simply means that the 420-nm photons with their 2.96-eV energy are not much above the frequency threshold. You can show for yourself that the threshold wavelength is 459 nm (blue light). This means that if calcium metal is used in a light meter, the meter will be insensitive to wavelengths longer than those of blue light. Such a light meter would be completely insensitive to red light, for example.

Note:

PhET Explorations: Photoelectric Effect

See how light knocks electrons off a metal target, and recreate the experiment that spawned the field of quantum mechanics.

<https://archive.cnx.org/specials/cf1152da-eae8-11e5-b874-f779884a9994/photoelectric-effect/#sim-photoelectric-effect>

Section Summary

- The photoelectric effect is the process in which EM radiation ejects electrons from a material.
- Einstein proposed photons to be quanta of EM radiation having energy $E = hf$, where f is the frequency of the radiation.

- All EM radiation is composed of photons. As Einstein explained, all characteristics of the photoelectric effect are due to the interaction of individual photons with individual electrons.
- The maximum kinetic energy KE_e of ejected electrons (photoelectrons) is given by $KE_e = hf - BE$, where hf is the photon energy and BE is the binding energy (or work function) of the electron to the particular material.

Conceptual Questions

Exercise:

Problem:

Is visible light the only type of EM radiation that can cause the photoelectric effect?

Exercise:

Problem:

Which aspects of the photoelectric effect cannot be explained without photons? Which can be explained without photons? Are the latter inconsistent with the existence of photons?

Exercise:

Problem:

Is the photoelectric effect a direct consequence of the wave character of EM radiation or of the particle character of EM radiation? Explain briefly.

Exercise:

Problem:

Insulators (nonmetals) have a higher BE than metals, and it is more difficult for photons to eject electrons from insulators. Discuss how this relates to the free charges in metals that make them good conductors.

Exercise:**Problem:**

If you pick up and shake a piece of metal that has electrons in it free to move as a current, no electrons fall out. Yet if you heat the metal, electrons can be boiled off. Explain both of these facts as they relate to the amount and distribution of energy involved with shaking the object as compared with heating it.

Problems & Exercises**Exercise:****Problem:**

What is the longest-wavelength EM radiation that can eject a photoelectron from silver, given that the binding energy is 4.73 eV? Is this in the visible range?

Solution:

263 nm

Exercise:**Problem:**

Find the longest-wavelength photon that can eject an electron from potassium, given that the binding energy is 2.24 eV. Is this visible EM radiation?

Exercise:**Problem:**

What is the binding energy in eV of electrons in magnesium, if the longest-wavelength photon that can eject electrons is 337 nm?

Solution:

3.69 eV

Exercise:

Problem:

Calculate the binding energy in eV of electrons in aluminum, if the longest-wavelength photon that can eject them is 304 nm.

Exercise:

Problem:

What is the maximum kinetic energy in eV of electrons ejected from sodium metal by 450-nm EM radiation, given that the binding energy is 2.28 eV?

Solution:

0.483 eV

Exercise:

Problem:

UV radiation having a wavelength of 120 nm falls on gold metal, to which electrons are bound by 4.82 eV. What is the maximum kinetic energy of the ejected photoelectrons?

Exercise:

Problem:

Violet light of wavelength 400 nm ejects electrons with a maximum kinetic energy of 0.860 eV from sodium metal. What is the binding energy of electrons to sodium metal?

Solution:

2.25 eV

Exercise:

Problem:

UV radiation having a 300-nm wavelength falls on uranium metal, ejecting 0.500-eV electrons. What is the binding energy of electrons to uranium metal?

Exercise:**Problem:**

What is the wavelength of EM radiation that ejects 2.00-eV electrons from calcium metal, given that the binding energy is 2.71 eV? What type of EM radiation is this?

Solution:

- (a) 264 nm
- (b) Ultraviolet

Exercise:**Problem:**

Find the wavelength of photons that eject 0.100-eV electrons from potassium, given that the binding energy is 2.24 eV. Are these photons visible?

Exercise:**Problem:**

What is the maximum velocity of electrons ejected from a material by 80-nm photons, if they are bound to the material by 4.73 eV?

Solution:

$$1.95 \times 10^6 \text{ m/s}$$

Exercise:

Problem:

Photoelectrons from a material with a binding energy of 2.71 eV are ejected by 420-nm photons. Once ejected, how long does it take these electrons to travel 2.50 cm to a detection device?

Exercise:**Problem:**

A laser with a power output of 2.00 mW at a wavelength of 400 nm is projected onto calcium metal. (a) How many electrons per second are ejected? (b) What power is carried away by the electrons, given that the binding energy is 2.71 eV?

Solution:

(a) $4.02 \times 10^{15} \text{ /s}$

(b) 0.256 mW

Exercise:**Problem:**

(a) Calculate the number of photoelectrons per second ejected from a 1.00-mm^2 area of sodium metal by 500-nm EM radiation having an intensity of 1.30 kW/m^2 (the intensity of sunlight above the Earth's atmosphere). (b) Given that the binding energy is 2.28 eV, what power is carried away by the electrons? (c) The electrons carry away less power than brought in by the photons. Where does the other power go? How can it be recovered?

Exercise:**Problem: Unreasonable Results**

Red light having a wavelength of 700 nm is projected onto magnesium metal to which electrons are bound by 3.68 eV. (a) Use $KE_e = hf - BE$ to calculate the kinetic energy of the ejected electrons.

(b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) -1.90 eV

(b) Negative kinetic energy

(c) That the electrons would be knocked free.

Exercise:

Problem: Unreasonable Results

(a) What is the binding energy of electrons to a material from which 4.00-eV electrons are ejected by 400-nm EM radiation? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

photoelectric effect

the phenomenon whereby some materials eject electrons when light is shined on them

photon

a quantum, or particle, of electromagnetic radiation

photon energy

the amount of energy a photon has; $E = hf$

binding energy

also called the *work function*; the amount of energy necessary to eject an electron from a material

Photon Energies and the Electromagnetic Spectrum

- Explain the relationship between the energy of a photon in joules or electron volts and its wavelength or frequency.
- Calculate the number of photons per second emitted by a monochromatic source of specific wavelength and power.

Ionizing Radiation

A photon is a quantum of EM radiation. Its energy is given by $E = hf$ and is related to the frequency f and wavelength λ of the radiation by

Equation:

$$E = hf = \frac{hc}{\lambda} (\text{energy of a photon}),$$

where E is the energy of a single photon and c is the speed of light. When working with small systems, energy in eV is often useful. Note that Planck's constant in these units is

Equation:

$$h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}.$$

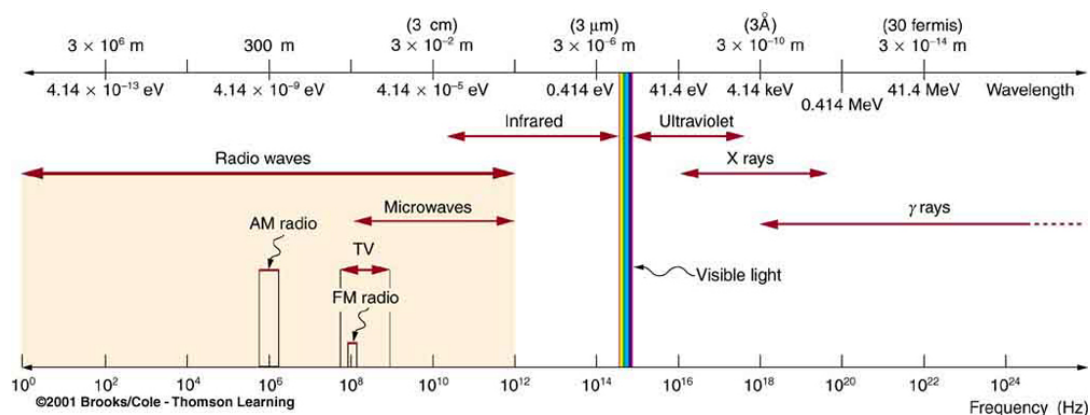
Since many wavelengths are stated in nanometers (nm), it is also useful to know that

Equation:

$$hc = 1240 \text{ eV} \cdot \text{nm}.$$

These will make many calculations a little easier.

All EM radiation is composed of photons. [\[link\]](#) shows various divisions of the EM spectrum plotted against wavelength, frequency, and photon energy. Previously in this book, photon characteristics were alluded to in the discussion of some of the characteristics of UV, x rays, and γ rays, the first of which start with frequencies just above violet in the visible spectrum. It was noted that these types of EM radiation have characteristics much different than visible light. We can now see that such properties arise because photon energy is larger at high frequencies.



The EM spectrum, showing major categories as a function of photon energy in eV, as well as wavelength and frequency. Certain characteristics of EM radiation are directly attributable to photon energy alone.

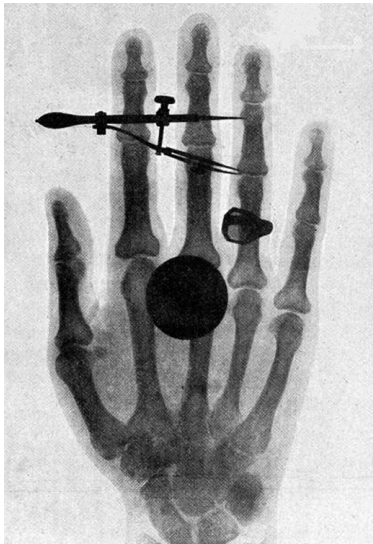
Rotational energies of molecules	10^{-5} eV
Vibrational energies of molecules	0.1 eV
Energy between outer electron shells in atoms	1 eV
Binding energy of a weakly bound molecule	1 eV
Energy of red light	2 eV
Binding energy of a tightly bound molecule	10 eV
Energy to ionize atom or molecule	10 to 1000 eV

Representative Energies for Submicroscopic Effects (Order of Magnitude Only)

Photons act as individual quanta and interact with individual electrons, atoms, molecules, and so on. The energy a photon carries is, thus, crucial to the effects it has. [\[link\]](#) lists representative submicroscopic energies in eV. When we compare photon energies from the EM spectrum in [\[link\]](#) with energies in the table, we can see how effects vary with the type of EM radiation.

Gamma rays, a form of nuclear and cosmic EM radiation, can have the highest frequencies and, hence, the highest photon energies in the EM spectrum. For example, a γ -ray photon with $f = 10^{21}$ Hz has an energy $E = hf = 6.63 \times 10^{-13} \text{ J} = 4.14 \text{ MeV}$. This is sufficient energy to ionize thousands of atoms and molecules, since only 10 to 1000 eV are needed per ionization. In fact, γ rays are one type of **ionizing radiation**, as are x rays and UV, because they produce ionization in materials that absorb

them. Because so much ionization can be produced, a single γ -ray photon can cause significant damage to biological tissue, killing cells or damaging their ability to properly reproduce. When cell reproduction is disrupted, the result can be cancer, one of the known effects of exposure to ionizing radiation. Since cancer cells are rapidly reproducing, they are exceptionally sensitive to the disruption produced by ionizing radiation. This means that ionizing radiation has positive uses in cancer treatment as well as risks in producing cancer.



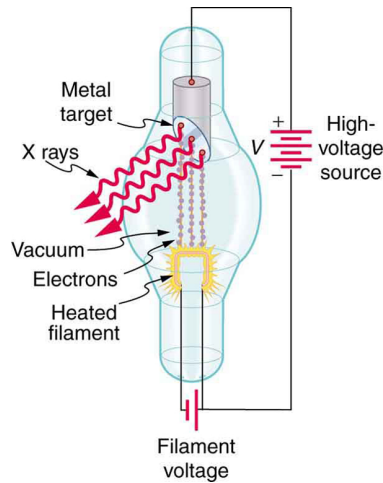
One of the first x-ray images, taken by Röntgen himself. The hand belongs to Bertha Röntgen, his wife. (credit: Wilhelm Conrad Röntgen, via Wikimedia Commons)

High photon energy also enables γ rays to penetrate materials, since a collision with a single atom or molecule is unlikely to absorb all the γ ray's energy. This can make γ rays useful as a probe, and they are sometimes used in medical imaging. **x rays**, as you can see in [\[link\]](#), overlap with the low-frequency end of the γ ray range. Since x rays have energies of keV and up, individual x-ray photons also can produce large amounts of ionization. At lower photon energies, x rays are not as penetrating as γ rays and are slightly less hazardous. X rays are ideal for medical imaging, their most common use, and a fact that was recognized immediately upon their discovery in 1895 by the German physicist W. C. Roentgen (1845–1923). (See [\[link\]](#).) Within one year of their discovery, x rays (for a time called Roentgen rays) were used for medical diagnostics. Roentgen received the 1901 Nobel Prize for the discovery of x rays.

Note:

Connections: Conservation of Energy

Once again, we find that conservation of energy allows us to consider the initial and final forms that energy takes, without having to make detailed calculations of the intermediate steps. [\[link\]](#) is solved by considering only the initial and final forms of energy.



X rays are produced when energetic electrons strike the copper anode of this cathode ray tube (CRT). Electrons (shown here as separate particles) interact individually with the material they strike, sometimes producing photons of EM radiation.

While γ rays originate in nuclear decay, x rays are produced by the process shown in [\[link\]](#). Electrons ejected by thermal agitation from a hot filament in a vacuum tube are accelerated through a high voltage, gaining kinetic energy from the electrical potential energy. When they strike the anode, the electrons convert their kinetic energy to a variety of forms, including thermal energy. But since an accelerated charge radiates EM waves, and since the electrons act individually, photons are also produced. Some of these x-ray photons obtain the kinetic energy of the electron. The accelerated electrons originate at the cathode, so such a tube is called a cathode ray tube (CRT), and various versions of them are found in older TV and computer screens as well as in x-ray machines.

Example:

X-ray Photon Energy and X-ray Tube Voltage

Find the maximum energy in eV of an x-ray photon produced by electrons accelerated through a potential difference of 50.0 kV in a CRT like the one in [\[link\]](#).

Strategy

Electrons can give all of their kinetic energy to a single photon when they strike the anode of a CRT. (This is something like the photoelectric effect in reverse.) The kinetic energy of the electron comes from electrical potential energy. Thus we can simply equate the maximum photon energy to the electrical potential energy—that is, $hf = qV$. (We do not have to calculate each step from beginning to end if we know that all of the starting energy qV is converted to the final form hf .)

Solution

The maximum photon energy is $hf = qV$, where q is the charge of the electron and V is the accelerating voltage. Thus,

Equation:

$$hf = (1.60 \times 10^{-19} \text{ C})(50.0 \times 10^3 \text{ V}).$$

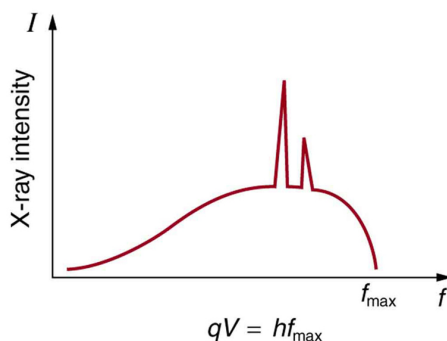
From the definition of the electron volt, we know $1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$, where $1 \text{ J} = 1 \text{ C} \cdot \text{V}$. Gathering factors and converting energy to eV yields

Equation:

$$hf = (50.0 \times 10^3)(1.60 \times 10^{-19} \text{ C} \cdot \text{V}) \left(\frac{1 \text{ eV}}{1.60 \times 10^{-19} \text{ C} \cdot \text{V}} \right) = (50.0 \times 10^3)(1 \text{ eV}) = 50.0 \text{ keV}.$$

Discussion

This example produces a result that can be applied to many similar situations. If you accelerate a single elementary charge, like that of an electron, through a potential given in volts, then its energy in eV has the same numerical value. Thus a 50.0-kV potential generates 50.0 keV electrons, which in turn can produce photons with a maximum energy of 50 keV. Similarly, a 100-kV potential in an x-ray tube can generate up to 100-keV x-ray photons. Many x-ray tubes have adjustable voltages so that various energy x rays with differing energies, and therefore differing abilities to penetrate, can be generated.



X-ray spectrum obtained when energetic electrons strike a material. The smooth part of the spectrum is bremsstrahlung, while the peaks are characteristic of the anode material. Both are atomic processes that produce energetic

photons known as x-ray
photons.

[\[link\]](#) shows the spectrum of x rays obtained from an x-ray tube. There are two distinct features to the spectrum. First, the smooth distribution results from electrons being decelerated in the anode material. A curve like this is obtained by detecting many photons, and it is apparent that the maximum energy is unlikely. This decelerating process produces radiation that is called **bremsstrahlung** (German for *braking radiation*). The second feature is the existence of sharp peaks in the spectrum; these are called **characteristic x rays**, since they are characteristic of the anode material. Characteristic x rays come from atomic excitations unique to a given type of anode material. They are akin to lines in atomic spectra, implying the energy levels of atoms are quantized. Phenomena such as discrete atomic spectra and characteristic x rays are explored further in [Atomic Physics](#).

Ultraviolet radiation (approximately 4 eV to 300 eV) overlaps with the low end of the energy range of x rays, but UV is typically lower in energy. UV comes from the de-excitation of atoms that may be part of a hot solid or gas. These atoms can be given energy that they later release as UV by numerous processes, including electric discharge, nuclear explosion, thermal agitation, and exposure to x rays. A UV photon has sufficient energy to ionize atoms and molecules, which makes its effects different from those of visible light. UV thus has some of the same biological effects as γ rays and x rays. For example, it can cause skin cancer and is used as a sterilizer. The major difference is that several UV photons are required to disrupt cell reproduction or kill a bacterium, whereas single γ -ray and X-ray photons can do the same damage. But since UV does have the energy to alter molecules, it can do what visible light cannot. One of the beneficial aspects of UV is that it triggers the production of vitamin D in the skin, whereas visible light has insufficient energy per photon to alter the molecules that trigger this production. Infantile jaundice is treated by exposing the baby to UV (with eye protection), called phototherapy, the beneficial effects of which are thought to be related to its ability to help prevent the buildup of potentially toxic bilirubin in the blood.

Example:

Photon Energy and Effects for UV

Short-wavelength UV is sometimes called vacuum UV, because it is strongly absorbed by air and must be studied in a vacuum. Calculate the photon energy in eV for 100-nm vacuum UV, and estimate the number of molecules it could ionize or break apart.

Strategy

Using the equation $E = hf$ and appropriate constants, we can find the photon energy and compare it with energy information in [\[link\]](#).

Solution

The energy of a photon is given by

Equation:

$$E = hf = \frac{hc}{\lambda}.$$

Using $hc = 1240 \text{ eV} \cdot \text{nm}$, we find that

Equation:

$$E = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{100 \text{ nm}} = 12.4 \text{ eV}.$$

Discussion

According to [\[link\]](#), this photon energy might be able to ionize an atom or molecule, and it is about what is needed to break up a tightly bound molecule, since they are bound by approximately 10 eV. This photon energy could destroy about a dozen weakly bound molecules. Because of its high photon energy, UV disrupts atoms and molecules it interacts with. One good consequence is that all but the longest-wavelength UV is strongly absorbed and is easily blocked by sunglasses. In fact, most of the Sun's UV is absorbed by a thin layer of ozone in the upper atmosphere, protecting sensitive organisms on Earth. Damage to our ozone layer by the addition of such chemicals as CFC's has reduced this protection for us.

Visible Light

The range of photon energies for **visible light** from red to violet is 1.63 to 3.26 eV, respectively (left for this chapter's Problems and Exercises to verify). These energies are on the order of those between outer electron shells in atoms and molecules. This means that these photons can be absorbed by atoms and molecules. A *single* photon can actually stimulate the retina, for example, by altering a receptor molecule that then triggers a nerve impulse. Photons can be absorbed or emitted only by atoms and molecules that have precisely the correct quantized energy step to do so. For example, if a red photon of frequency f encounters a molecule that has an energy step, ΔE , equal to hf , then the photon can be absorbed. Violet flowers absorb red and reflect violet; this implies there is no energy step between levels in the receptor molecule equal to the violet photon's energy, but there is an energy step for the red.

There are some noticeable differences in the characteristics of light between the two ends of the visible spectrum that are due to photon energies. Red light has insufficient photon energy to expose most black-and-white film, and it is thus used to illuminate darkrooms where such film is developed. Since violet light has a higher photon energy, dyes that absorb violet tend to fade more quickly than those that do not. (See [\[link\]](#).) Take a look at some faded color posters in a storefront some time, and you will notice that the blues and violets are the last to fade. This is because other dyes, such as red and green dyes, absorb blue and violet photons, the higher energies of which break up their weakly bound molecules. (Complex molecules such as those in dyes and DNA tend to be weakly bound.) Blue and violet dyes reflect those colors and, therefore, do not absorb these more energetic photons, thus suffering less molecular damage.



Why do the reds, yellows,
and greens fade before
the blues and violets
when exposed to the Sun,
as with this poster? The
answer is related to
photon energy. (credit:
Deb Collins, Flickr)

Transparent materials, such as some glasses, do not absorb any visible light, because there is no energy step in the atoms or molecules that could absorb the light. Since individual photons interact with individual atoms, it is nearly impossible to have two photons absorbed simultaneously to reach a large energy step. Because of its lower photon energy, visible light can sometimes pass through many kilometers of a substance, while higher frequencies like UV, x ray, and γ rays are absorbed, because they have sufficient photon energy to ionize the material.

Example:

How Many Photons per Second Does a Typical Light Bulb Produce?

Assuming that 10.0% of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, calculate the number of visible photons emitted per second.

Strategy

Power is energy per unit time, and so if we can find the energy per photon, we can determine the number of photons per second. This will best be done in joules, since power is given in watts, which are joules per second.

Solution

The power in visible light production is 10.0% of 100 W, or 10.0 J/s. The energy of the average visible photon is found by substituting the given average wavelength into the formula

Equation:

$$E = \frac{hc}{\lambda}.$$

This produces

Equation:

$$E = \frac{(6.63 \times 10^{-34} \text{ J} \cdot \text{s})(3.00 \times 10^8 \text{ m/s})}{580 \times 10^{-9} \text{ m}} = 3.43 \times 10^{-19} \text{ J}.$$

The number of visible photons per second is thus

Equation:

$$\text{photon/s} = \frac{10.0 \text{ J/s}}{3.43 \times 10^{-19} \text{ J/photon}} = 2.92 \times 10^{19} \text{ photon/s}.$$

Discussion

This incredible number of photons per second is verification that individual photons are insignificant in ordinary human experience. It is also a verification of the correspondence principle—on the macroscopic scale, quantization becomes essentially continuous or classical. Finally, there are so many photons emitted by a 100-W lightbulb that it can be seen by the unaided eye many kilometers away.

Lower-Energy Photons

Infrared radiation (IR) has even lower photon energies than visible light and cannot significantly alter atoms and molecules. IR can be absorbed and emitted by atoms and molecules, particularly between closely spaced states. IR is extremely strongly absorbed by water, for example, because water molecules have many states separated by energies on the order of 10^{-5} eV to 10^{-2} eV, well within the IR and microwave energy ranges. This is why in the IR range, skin is almost jet black, with an emissivity near 1—there are many states in water molecules in the skin that can absorb a large range of IR photon energies. Not all molecules have this property. Air, for example, is nearly transparent to many IR frequencies.

Microwaves are the highest frequencies that can be produced by electronic circuits, although they are also produced naturally. Thus microwaves are similar to IR but do not extend to as high frequencies. There are states in water and other molecules that have the same frequency and energy as microwaves, typically about 10^{-5} eV. This is one reason why food absorbs microwaves more strongly than many other materials, making microwave ovens an efficient way of putting energy directly into food.

Photon energies for both IR and microwaves are so low that huge numbers of photons are involved in any significant energy transfer by IR or microwaves (such as warming yourself with a heat lamp or cooking pizza in the microwave). Visible light, IR, microwaves, and all lower frequencies cannot produce ionization with single photons and do not ordinarily have the hazards of higher frequencies. When visible, IR, or microwave radiation is hazardous, such as the inducement of cataracts by microwaves, the hazard is due to huge numbers of photons acting together (not to an accumulation of photons, such as sterilization by weak UV). The negative effects of visible, IR, or microwave radiation can be thermal effects, which could be produced by any heat source. But one difference is that at very high intensity, strong electric and magnetic fields can be produced by photons acting together. Such electromagnetic fields (EMF) can actually ionize materials.

Note:**Misconception Alert: High-Voltage Power Lines**

Although some people think that living near high-voltage power lines is hazardous to one's health, ongoing studies of the transient field effects produced by these lines show their strengths to be insufficient to cause damage. Demographic studies also fail to show significant correlation of ill effects with high-voltage power lines. The American Physical Society issued a report over 10 years ago on power-line fields, which concluded that the scientific literature and reviews of panels show no consistent, significant link between cancer and power-line fields. They also felt that the "diversion of resources to eliminate a threat which has no persuasive scientific basis is disturbing."

It is virtually impossible to detect individual photons having frequencies below microwave frequencies, because of their low photon energy. But the photons are there. A continuous EM wave can be modeled as photons. At low frequencies, EM waves are generally treated as time- and position-varying electric and magnetic fields with no discernible quantization. This is another example of the correspondence principle in situations involving huge numbers of photons.

Note:**PhET Explorations: Color Vision**

Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

https://phet.colorado.edu/sims/html/color-vision/latest/color-vision_en.html

Section Summary

- Photon energy is responsible for many characteristics of EM radiation, being particularly noticeable at high frequencies.
- Photons have both wave and particle characteristics.

Conceptual Questions**Exercise:**

Problem: Why are UV, x rays, and γ rays called ionizing radiation?

Exercise:**Problem:**

How can treating food with ionizing radiation help keep it from spoiling? UV is not very penetrating. What else could be used?

Exercise:

Problem:

Some television tubes are CRTs. They use an approximately 30-kV accelerating potential to send electrons to the screen, where the electrons stimulate phosphors to emit the light that forms the pictures we watch. Would you expect x rays also to be created?

Exercise:**Problem:**

Tanning salons use “safe” UV with a longer wavelength than some of the UV in sunlight. This “safe” UV has enough photon energy to trigger the tanning mechanism. Is it likely to be able to cause cell damage and induce cancer with prolonged exposure?

Exercise:**Problem:**

Your pupils dilate when visible light intensity is reduced. Does wearing sunglasses that lack UV blockers increase or decrease the UV hazard to your eyes? Explain.

Exercise:**Problem:**

One could feel heat transfer in the form of infrared radiation from a large nuclear bomb detonated in the atmosphere 75 km from you. However, none of the profusely emitted x rays or γ rays reaches you. Explain.

Exercise:

Problem: Can a single microwave photon cause cell damage? Explain.

Exercise:**Problem:**

In an x-ray tube, the maximum photon energy is given by $hf = qV$. Would it be technically more correct to say $hf = qV + BE$, where BE is the binding energy of electrons in the target anode? Why isn't the energy stated the latter way?

Problems & Exercises**Exercise:****Problem:**

What is the energy in joules and eV of a photon in a radio wave from an AM station that has a 1530-kHz broadcast frequency?

Solution:

6.34×10^{-9} eV, 1.01×10^{-27} J

Exercise:

Problem:

(a) Find the energy in joules and eV of photons in radio waves from an FM station that has a 90.0-MHz broadcast frequency. (b) What does this imply about the number of photons per second that the radio station must broadcast?

Exercise:

Problem: Calculate the frequency in hertz of a 1.00-MeV γ -ray photon.

Solution:

$$2.42 \times 10^{20} \text{ Hz}$$

Exercise:**Problem:**

(a) What is the wavelength of a 1.00-eV photon? (b) Find its frequency in hertz. (c) Identify the type of EM radiation.

Exercise:**Problem:**

Do the unit conversions necessary to show that $hc = 1240 \text{ eV} \cdot \text{nm}$, as stated in the text.

Solution:**Equation:**

$$\begin{aligned} hc &= (6.62607 \times 10^{-34} \text{ J} \cdot \text{s}) (2.99792 \times 10^8 \text{ m/s}) \left(\frac{10^9 \text{ nm}}{1 \text{ m}} \right) \left(\frac{1.00000 \text{ eV}}{1.60218 \times 10^{-19} \text{ J}} \right) \\ &= 1239.84 \text{ eV} \cdot \text{nm} \\ &\approx 1240 \text{ eV} \cdot \text{nm} \end{aligned}$$

Exercise:**Problem:**

Confirm the statement in the text that the range of photon energies for visible light is 1.63 to 3.26 eV, given that the range of visible wavelengths is 380 to 760 nm.

Exercise:**Problem:**

(a) Calculate the energy in eV of an IR photon of frequency $2.00 \times 10^{13} \text{ Hz}$. (b) How many of these photons would need to be absorbed simultaneously by a tightly bound molecule to break it apart? (c) What is the energy in eV of a γ ray of frequency $3.00 \times 10^{20} \text{ Hz}$? (d) How many tightly bound molecules could a single such γ ray break apart?

Solution:

(a) 0.0829 eV

- (b) 121
- (c) 1.24 MeV
- (d) 1.24×10^5

Exercise:

Problem: Prove that, to three-digit accuracy, $h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}$, as stated in the text.

Exercise:

Problem:

(a) What is the maximum energy in eV of photons produced in a CRT using a 25.0-kV accelerating potential, such as a color TV? (b) What is their frequency?

Solution:

- (a) $25.0 \times 10^3 \text{ eV}$
- (b) $6.04 \times 10^{18} \text{ Hz}$

Exercise:

Problem:

What is the accelerating voltage of an x-ray tube that produces x rays with a shortest wavelength of 0.0103 nm?

Exercise:

Problem:

(a) What is the ratio of power outputs by two microwave ovens having frequencies of 950 and 2560 MHz, if they emit the same number of photons per second? (b) What is the ratio of photons per second if they have the same power output?

Solution:

- (a) 2.69
- (b) 0.371

Exercise:

Problem:

How many photons per second are emitted by the antenna of a microwave oven, if its power output is 1.00 kW at a frequency of 2560 MHz?

Exercise:

Problem:

Some satellites use nuclear power. (a) If such a satellite emits a 1.00-W flux of γ rays having an average energy of 0.500 MeV, how many are emitted per second? (b) These γ rays affect other satellites. How far away must another satellite be to only receive one γ ray per second per square meter?

Solution:

(a) 1.25×10^{13} photons/s

(b) 997 km

Exercise:**Problem:**

(a) If the power output of a 650-kHz radio station is 50.0 kW, how many photons per second are produced? (b) If the radio waves are broadcast uniformly in all directions, find the number of photons per second per square meter at a distance of 100 km. Assume no reflection from the ground or absorption by the air.

Exercise:**Problem:**

How many x-ray photons per second are created by an x-ray tube that produces a flux of x rays having a power of 1.00 W? Assume the average energy per photon is 75.0 keV.

Solution:

8.33×10^{13} photons/s

Exercise:**Problem:**

(a) How far away must you be from a 650-kHz radio station with power 50.0 kW for there to be only one photon per second per square meter? Assume no reflections or absorption, as if you were in deep outer space. (b) Discuss the implications for detecting intelligent life in other solar systems by detecting their radio broadcasts.

Exercise:**Problem:**

Assuming that 10.0% of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, and that the photons spread out uniformly and are not absorbed by the atmosphere, how far away would you be if 500 photons per second enter the 3.00-mm diameter pupil of your eye? (This number easily stimulates the retina.)

Solution:

181 km

Exercise:

Problem:Construct Your Own Problem

Consider a laser pen. Construct a problem in which you calculate the number of photons per second emitted by the pen. Among the things to be considered are the laser pen's wavelength and power output. Your instructor may also wish for you to determine the minimum diffraction spreading in the beam and the number of photons per square centimeter the pen can project at some large distance. In this latter case, you will also need to consider the output size of the laser beam, the distance to the object being illuminated, and any absorption or scattering along the way.

Glossary

gamma ray

also γ -ray; highest-energy photon in the EM spectrum

ionizing radiation

radiation that ionizes materials that absorb it

x ray

EM photon between γ -ray and UV in energy

bremsstrahlung

German for *braking radiation*; produced when electrons are decelerated

characteristic x rays

x rays whose energy depends on the material they were produced in

ultraviolet radiation

UV; ionizing photons slightly more energetic than violet light

visible light

the range of photon energies the human eye can detect

infrared radiation

photons with energies slightly less than red light

microwaves

photons with wavelengths on the order of a micron (μm)

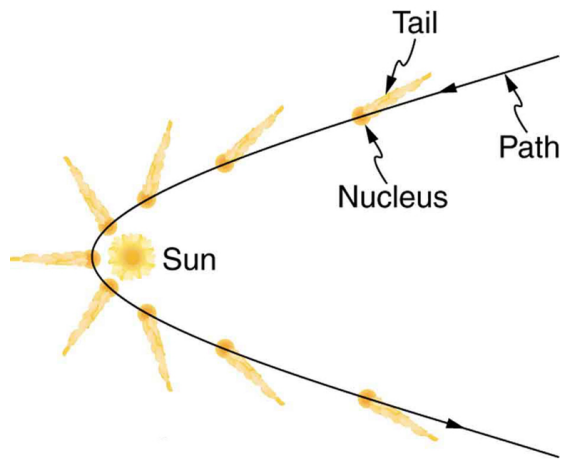
Photon Momentum

- Relate the linear momentum of a photon to its energy or wavelength, and apply linear momentum conservation to simple processes involving the emission, absorption, or reflection of photons.
- Account qualitatively for the increase of photon wavelength that is observed, and explain the significance of the Compton wavelength.

Measuring Photon Momentum

The quantum of EM radiation we call a **photon** has properties analogous to those of particles we can see, such as grains of sand. A photon interacts as a unit in collisions or when absorbed, rather than as an extensive wave.

Massive quanta, like electrons, also act like macroscopic particles—something we expect, because they are the smallest units of matter. Particles carry momentum as well as energy. Despite photons having no mass, there has long been evidence that EM radiation carries momentum. (Maxwell and others who studied EM waves predicted that they would carry momentum.) It is now a well-established fact that photons *do* have momentum. In fact, photon momentum is suggested by the photoelectric effect, where photons knock electrons out of a substance. [\[link\]](#) shows macroscopic evidence of photon momentum.



The tails of the Hale-Bopp comet point away from the Sun, evidence that light has momentum. Dust emanating from the body of the comet forms this tail. Particles of dust are pushed away from the Sun by light reflecting from them. The blue ionized gas tail is also produced by photons interacting with atoms in the comet material. (credit: Geoff Chester, U.S. Navy, via Wikimedia Commons)

[\[link\]](#) shows a comet with two prominent tails. What most people do not know about the tails is that they always point *away* from the Sun rather than trailing behind the comet (like the tail of Bo Peep's sheep). Comet tails are composed of gases and dust evaporated from the body of the comet and ionized gas. The dust particles recoil away from the Sun when photons scatter from them. Evidently, photons carry momentum in the direction of their motion (away from the Sun), and some of this momentum is transferred to dust particles in collisions. Gas atoms and molecules in the blue tail are most affected by other particles of radiation, such as protons and electrons emanating from the Sun, rather than by the momentum of photons.

Note:

Connections: Conservation of Momentum

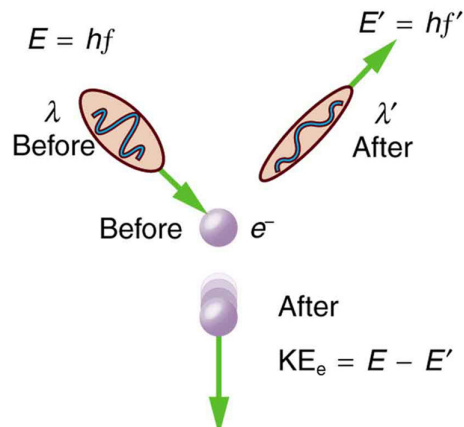
Not only is momentum conserved in all realms of physics, but all types of particles are found to have momentum. We expect particles with mass to have momentum, but now we see that massless particles including photons also carry momentum.

Momentum is conserved in quantum mechanics just as it is in relativity and classical physics. Some of the earliest direct experimental evidence of this came from scattering of x-ray photons by electrons in substances, named Compton scattering after the American physicist Arthur H. Compton (1892–1962). Around 1923, Compton observed that x rays scattered from materials had a decreased energy and correctly analyzed this as being due to the scattering of photons from electrons. This phenomenon could be handled as a collision between two particles—a photon and an electron at rest in the material. Energy and momentum are conserved in the collision. (See [\[link\]](#)) He won a Nobel Prize in 1929 for the discovery of this scattering, now called the **Compton effect**, because it helped prove that **photon momentum** is given by

Equation:

$$p = \frac{h}{\lambda},$$

where h is Planck's constant and λ is the photon wavelength. (Note that relativistic momentum given as $p = \gamma mu$ is valid only for particles having mass.)



The Compton effect is the name given to the scattering of a photon by an electron. Energy and momentum are conserved, resulting in a reduction of both for the scattered photon. Studying this effect, Compton verified that photons have momentum.

We can see that photon momentum is small, since $p = h/\lambda$ and h is very small. It is for this reason that we do not ordinarily observe photon

momentum. Our mirrors do not recoil when light reflects from them (except perhaps in cartoons). Compton saw the effects of photon momentum because he was observing x rays, which have a small wavelength and a relatively large momentum, interacting with the lightest of particles, the electron.

Example:

Electron and Photon Momentum Compared

(a) Calculate the momentum of a visible photon that has a wavelength of 500 nm. (b) Find the velocity of an electron having the same momentum. (c) What is the energy of the electron, and how does it compare with the energy of the photon?

Strategy

Finding the photon momentum is a straightforward application of its definition: $p = \frac{h}{\lambda}$. If we find the photon momentum is small, then we can assume that an electron with the same momentum will be nonrelativistic, making it easy to find its velocity and kinetic energy from the classical formulas.

Solution for (a)

Photon momentum is given by the equation:

Equation:

$$p = \frac{h}{\lambda}.$$

Entering the given photon wavelength yields

Equation:

$$p = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{500 \times 10^{-9} \text{ m}} = 1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}.$$

Solution for (b)

Since this momentum is indeed small, we will use the classical expression $p = mv$ to find the velocity of an electron with this momentum. Solving for v and using the known value for the mass of an electron gives

Equation:

$$v = \frac{p}{m} = \frac{1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}}{9.11 \times 10^{-31} \text{ kg}} = 1460 \text{ m/s} \approx 1460 \text{ m/s}.$$

Solution for (c)

The electron has kinetic energy, which is classically given by

Equation:

$$\text{KE}_e = \frac{1}{2}mv^2.$$

Thus,

Equation:

$$\text{KE}_e = \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(1455 \text{ m/s})^2 = 9.64 \times 10^{-25} \text{ J}.$$

Converting this to eV by multiplying by $(1 \text{ eV})/(1.602 \times 10^{-19} \text{ J})$ yields

Equation:

$$\text{KE}_e = 6.02 \times 10^{-6} \text{ eV}.$$

The photon energy E is

Equation:

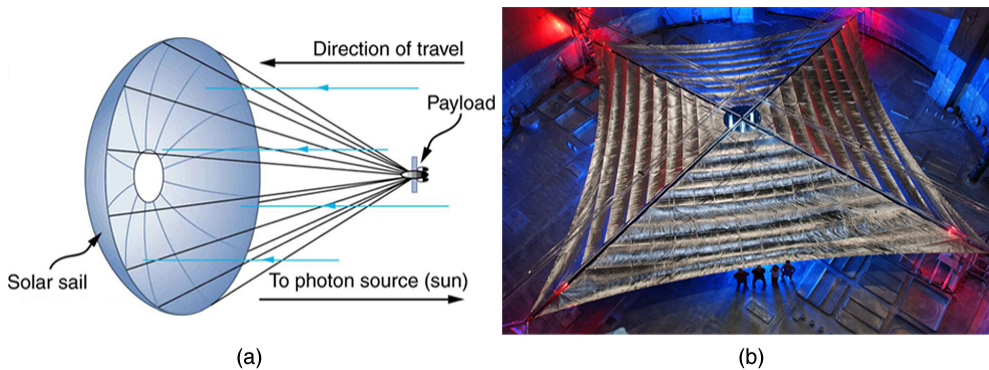
$$E = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{500 \text{ nm}} = 2.48 \text{ eV},$$

which is about five orders of magnitude greater.

Discussion

Photon momentum is indeed small. Even if we have huge numbers of them, the total momentum they carry is small. An electron with the same momentum has a 1460 m/s velocity, which is clearly nonrelativistic. A more massive particle with the same momentum would have an even smaller velocity. This is borne out by the fact that it takes far less energy to give an electron the same momentum as a photon. But on a quantum-mechanical scale, especially for high-energy photons interacting with small

masses, photon momentum is significant. Even on a large scale, photon momentum can have an effect if there are enough of them and if there is nothing to prevent the slow recoil of matter. Comet tails are one example, but there are also proposals to build space sails that use huge low-mass mirrors (made of aluminized Mylar) to reflect sunlight. In the vacuum of space, the mirrors would gradually recoil and could actually take spacecraft from place to place in the solar system. (See [\[link\]](#).)



(a) Space sails have been proposed that use the momentum of sunlight reflecting from gigantic low-mass sails to propel spacecraft about the solar system. A Russian test model of this (the Cosmos 1) was launched in 2005, but did not make it into orbit due to a rocket failure. (b) A U.S. version of this, labeled LightSail-1, is scheduled for trial launches in the first part of this decade. It will have a 40-m² sail. (credit: Kim Newton/NASA)

Relativistic Photon Momentum

There is a relationship between photon momentum p and photon energy E that is consistent with the relation given previously for the relativistic total energy of a particle as $E^2 = (pc)^2 + (mc)^2$. We know m is zero for a photon, but p is not, so that $E^2 = (pc)^2 + (mc)^2$ becomes

Equation:

$$E = pc,$$

or

Equation:

$$p = \frac{E}{c} \text{ (photons).}$$

To check the validity of this relation, note that $E = hc/\lambda$ for a photon. Substituting this into $p = E/c$ yields

Equation:

$$p = (hc/\lambda)/c = \frac{h}{\lambda},$$

as determined experimentally and discussed above. Thus, $p = E/c$ is equivalent to Compton's result $p = h/\lambda$. For a further verification of the relationship between photon energy and momentum, see [\[link\]](#).

Note:

Photon Detectors

Almost all detection systems talked about thus far—eyes, photographic plates, photomultiplier tubes in microscopes, and CCD cameras—rely on particle-like properties of photons interacting with a sensitive area. A change is caused and either the change is cascaded or zillions of points are recorded to form an image we detect. These detectors are used in biomedical imaging systems, and there is ongoing research into improving the efficiency of receiving photons, particularly by cooling detection systems and reducing thermal effects.

Example:**Photon Energy and Momentum**

Show that $p = E/c$ for the photon considered in the [\[link\]](#).

Strategy

We will take the energy E found in [\[link\]](#), divide it by the speed of light, and see if the same momentum is obtained as before.

Solution

Given that the energy of the photon is 2.48 eV and converting this to joules, we get

Equation:

$$p = \frac{E}{c} = \frac{(2.48 \text{ eV})(1.60 \times 10^{-19} \text{ J/eV})}{3.00 \times 10^8 \text{ m/s}} = 1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}.$$

Discussion

This value for momentum is the same as found before (note that unrounded values are used in all calculations to avoid even small rounding errors), an expected verification of the relationship $p = E/c$. This also means the relationship between energy, momentum, and mass given by $E^2 = (pc)^2 + (mc)^2$ applies to both matter and photons. Once again, note that p is not zero, even when m is.

Note:**Problem-Solving Suggestion**

Note that the forms of the constants $h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}$ and $hc = 1240 \text{ eV} \cdot \text{nm}$ may be particularly useful for this section's Problems and Exercises.

Section Summary

- Photons have momentum, given by $p = \frac{h}{\lambda}$, where λ is the photon wavelength.

- Photon energy and momentum are related by $p = \frac{E}{c}$, where $E = hf = hc/\lambda$ for a photon.

Conceptual Questions

Exercise:

Problem:

Which formula may be used for the momentum of all particles, with or without mass?

Exercise:

Problem:

Is there any measurable difference between the momentum of a photon and the momentum of matter?

Exercise:

Problem:

Why don't we feel the momentum of sunlight when we are on the beach?

Problems & Exercises

Exercise:

Problem:

- (a) Find the momentum of a 4.00-cm-wavelength microwave photon.
- (b) Discuss why you expect the answer to (a) to be very small.

Solution:

- (a) $1.66 \times 10^{-32} \text{ kg} \cdot \text{m/s}$

(b) The wavelength of microwave photons is large, so the momentum they carry is very small.

Exercise:

Problem:

(a) What is the momentum of a 0.0100-nm-wavelength photon that could detect details of an atom? (b) What is its energy in MeV?

Exercise:

Problem:

(a) What is the wavelength of a photon that has a momentum of $5.00 \times 10^{-29} \text{ kg} \cdot \text{m/s}$? (b) Find its energy in eV.

Solution:

(a) $13.3 \text{ } \mu\text{m}$

(b) $9.38 \times 10^{-2} \text{ eV}$

Exercise:

Problem:

(a) A γ -ray photon has a momentum of $8.00 \times 10^{-21} \text{ kg} \cdot \text{m/s}$. What is its wavelength? (b) Calculate its energy in MeV.

Exercise:

Problem:

(a) Calculate the momentum of a photon having a wavelength of $2.50 \text{ } \mu\text{m}$. (b) Find the velocity of an electron having the same momentum. (c) What is the kinetic energy of the electron, and how does it compare with that of the photon?

Solution:

(a) $2.65 \times 10^{-28} \text{ kg} \cdot \text{m/s}$

(b) 291 m/s

(c) electron 3.86×10^{-26} J, photon 7.96×10^{-20} J, ratio 2.06×10^6

Exercise:

Problem:

Repeat the previous problem for a 10.0-nm-wavelength photon.

Exercise:

Problem:

(a) Calculate the wavelength of a photon that has the same momentum as a proton moving at 1.00% of the speed of light. (b) What is the energy of the photon in MeV? (c) What is the kinetic energy of the proton in MeV?

Solution:

(a) 1.32×10^{-13} m

(b) 9.39 MeV

(c) 4.70×10^{-2} MeV

Exercise:

Problem:

(a) Find the momentum of a 100-keV x-ray photon. (b) Find the equivalent velocity of a neutron with the same momentum. (c) What is the neutron's kinetic energy in keV?

Exercise:

Problem:

Take the ratio of relativistic rest energy, $E = \gamma mc^2$, to relativistic momentum, $p = \gamma mu$, and show that in the limit that mass approaches zero, you find $E/p = c$.

Solution:

$E = \gamma mc^2$ and $P = \gamma mu$, so

Equation:

$$\frac{E}{P} = \frac{\gamma mc^2}{\gamma mu} = \frac{c^2}{u}.$$

As the mass of particle approaches zero, its velocity u will approach c , so that the ratio of energy to momentum in this limit is

Equation:

$$\lim_{m \rightarrow 0} \frac{E}{P} = \frac{c^2}{c} = c$$

which is consistent with the equation for photon energy.

Exercise:**Problem: Construct Your Own Problem**

Consider a space sail such as mentioned in [\[link\]](#). Construct a problem in which you calculate the light pressure on the sail in N/m^2 produced by reflecting sunlight. Also calculate the force that could be produced and how much effect that would have on a spacecraft. Among the things to be considered are the intensity of sunlight, its average wavelength, the number of photons per square meter this implies, the area of the space sail, and the mass of the system being accelerated.

Exercise:**Problem: Unreasonable Results**

A car feels a small force due to the light it sends out from its headlights, equal to the momentum of the light divided by the time in which it is emitted. (a) Calculate the power of each headlight, if they

exert a total force of 2.00×10^{-2} N backward on the car. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 3.00×10^6 W

(b) Headlights are way too bright.

(c) Force is too large.

Glossary

photon momentum

the amount of momentum a photon has, calculated by $p = \frac{h}{\lambda} = \frac{E}{c}$

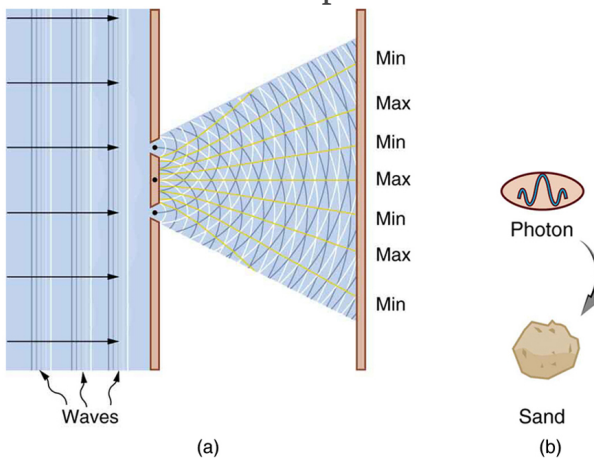
Compton effect

the phenomenon whereby x rays scattered from materials have decreased energy

The Particle-Wave Duality

- Explain what the term particle-wave duality means, and why it is applied to EM radiation.

We have long known that EM radiation is a wave, capable of interference and diffraction. We now see that light can be modeled as photons, which are massless particles. This may seem contradictory, since we ordinarily deal with large objects that never act like both wave and particle. An ocean wave, for example, looks nothing like a rock. To understand small-scale phenomena, we make analogies with the large-scale phenomena we observe directly. When we say something behaves like a wave, we mean it shows interference effects analogous to those seen in overlapping water waves. (See [\[link\]](#).) Two examples of waves are sound and EM radiation. When we say something behaves like a particle, we mean that it interacts as a discrete unit with no interference effects. Examples of particles include electrons, atoms, and photons of EM radiation. How do we talk about a phenomenon that acts like both a particle and a wave?



(a) The interference pattern for light through a double slit is a wave property understood by analogy to water waves. (b) The properties of photons having quantized energy and momentum and acting as a concentrated unit are

understood by analogy to
macroscopic particles.

There is no doubt that EM radiation interferes and has the properties of wavelength and frequency. There is also no doubt that it behaves as particles—photons with discrete energy. We call this twofold nature the **particle-wave duality**, meaning that EM radiation has both particle and wave properties. This so-called duality is simply a term for properties of the photon analogous to phenomena we can observe directly, on a macroscopic scale. If this term seems strange, it is because we do not ordinarily observe details on the quantum level directly, and our observations yield either particle *or* wavelike properties, but never both simultaneously.

Since we have a particle-wave duality for photons, and since we have seen connections between photons and matter in that both have momentum, it is reasonable to ask whether there is a particle-wave duality for matter as well. If the EM radiation we once thought to be a pure wave has particle properties, is it possible that matter has wave properties? The answer is yes. The consequences are tremendous, as we will begin to see in the next section.

Note:

PhET Explorations: Quantum Wave Interference

When do photons, electrons, and atoms behave like particles and when do they behave like waves? Watch waves spread out and interfere as they pass through a double slit, then get detected on a screen as tiny dots. Use quantum detectors to explore how measurements change the waves and the patterns they produce on the screen.

[Quantum](#)
[Wave](#)
[Interferenc](#)
[e](#)



Section Summary

- EM radiation can behave like either a particle or a wave.
- This is termed particle-wave duality.

Glossary

particle-wave duality

the property of behaving like either a particle or a wave; the term for the phenomenon that all particles have wave characteristics

The Wave Nature of Matter

- Describe the Davisson-Germer experiment, and explain how it provides evidence for the wave nature of electrons.

De Broglie Wavelength

In 1923 a French physics graduate student named Prince Louis-Victor de Broglie (1892–1987) made a radical proposal based on the hope that nature is symmetric. If EM radiation has both particle and wave properties, then nature would be symmetric if matter also had both particle and wave properties. If what we once thought of as an unequivocal wave (EM radiation) is also a particle, then what we think of as an unequivocal particle (matter) may also be a wave. De Broglie's suggestion, made as part of his doctoral thesis, was so radical that it was greeted with some skepticism. A copy of his thesis was sent to Einstein, who said it was not only probably correct, but that it might be of fundamental importance. With the support of Einstein and a few other prominent physicists, de Broglie was awarded his doctorate.

De Broglie took both relativity and quantum mechanics into account to develop the proposal that *all particles have a wavelength*, given by

Equation:

$$\lambda = \frac{h}{p} \text{ (matter and photons),}$$

where h is Planck's constant and p is momentum. This is defined to be the **de Broglie wavelength**. (Note that we already have this for photons, from the equation $p = h/\lambda$.) The hallmark of a wave is interference. If matter is a wave, then it must exhibit constructive and destructive interference. Why isn't this ordinarily observed? The answer is that in order to see significant interference effects, a wave must interact with an object about the same size as its wavelength. Since h is very small, λ is also small, especially for macroscopic objects. A 3-kg bowling ball moving at 10 m/s, for example, has

Equation:

$$\lambda = h/p = (6.63 \times 10^{-34} \text{ J}\cdot\text{s})/[(3 \text{ kg})(10 \text{ m/s})] = 2 \times 10^{-35} \text{ m}.$$

This means that to see its wave characteristics, the bowling ball would have to interact with something about 10^{-35} m in size—far smaller than anything known. When waves interact with objects much larger than their wavelength, they show negligible interference effects and move in straight lines (such as light rays in geometric optics). To get easily observed interference effects from particles of matter, the longest wavelength and hence smallest mass possible would be useful. Therefore, this effect was first observed with electrons.

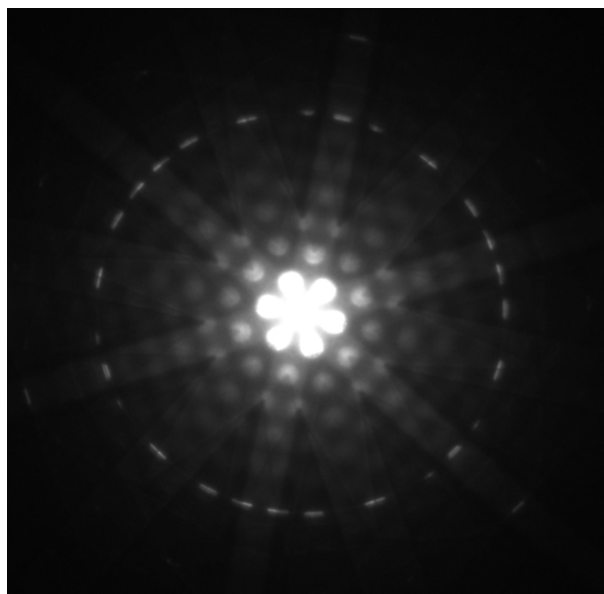
American physicists Clinton J. Davisson and Lester H. Germer in 1925 and, independently, British physicist G. P. Thomson (son of J. J. Thomson, discoverer of the electron) in 1926 scattered electrons from crystals and found diffraction patterns. These patterns are exactly consistent with interference of electrons having the de Broglie wavelength and are somewhat analogous to light interacting with a diffraction grating. (See [\[link\]](#).)

Note:**Connections: Waves**

All microscopic particles, whether massless, like photons, or having mass, like electrons, have wave properties. The relationship between momentum and wavelength is fundamental for all particles.

De Broglie's proposal of a wave nature for all particles initiated a remarkably productive era in which the foundations for quantum mechanics were laid. In 1926, the Austrian physicist Erwin Schrödinger (1887–1961) published four papers in which the wave nature of particles was treated explicitly with wave equations. At the same time, many others began important work. Among them was German physicist Werner Heisenberg

(1901–1976) who, among many other contributions to quantum mechanics, formulated a mathematical treatment of the wave nature of matter that used matrices rather than wave equations. We will deal with some specifics in later sections, but it is worth noting that de Broglie's work was a watershed for the development of quantum mechanics. De Broglie was awarded the Nobel Prize in 1929 for his vision, as were Davisson and G. P. Thomson in 1937 for their experimental verification of de Broglie's hypothesis.



This diffraction pattern was obtained for electrons diffracted by crystalline silicon. Bright regions are those of constructive interference, while dark regions are those of destructive interference. (credit: Ndtthe, Wikimedia Commons)

Example:

Electron Wavelength versus Velocity and Energy

For an electron having a de Broglie wavelength of 0.167 nm (appropriate for interacting with crystal lattice structures that are about this size): (a)

Calculate the electron's velocity, assuming it is nonrelativistic. (b)

Calculate the electron's kinetic energy in eV.

Strategy

For part (a), since the de Broglie wavelength is given, the electron's velocity can be obtained from $\lambda = h/p$ by using the nonrelativistic formula for momentum, $p = mv$. For part (b), once v is obtained (and it has been verified that v is nonrelativistic), the classical kinetic energy is simply $(1/2)mv^2$.

Solution for (a)

Substituting the nonrelativistic formula for momentum ($p = mv$) into the de Broglie wavelength gives

Equation:

$$\lambda = \frac{h}{p} = \frac{h}{mv}.$$

Solving for v gives

Equation:

$$v = \frac{h}{m\lambda}.$$

Substituting known values yields

Equation:

$$v = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{(9.11 \times 10^{-31} \text{ kg})(0.167 \times 10^{-9} \text{ m})} = 4.36 \times 10^6 \text{ m/s}.$$

Solution for (b)

While fast compared with a car, this electron's speed is not highly relativistic, and so we can comfortably use the classical formula to find the electron's kinetic energy and convert it to eV as requested.

Equation:

$$\begin{aligned}
 \text{KE} &= \frac{1}{2}mv^2 \\
 &= \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(4.36 \times 10^6 \text{ m/s})^2 \\
 &= (86.4 \times 10^{-18} \text{ J}) \left(\frac{1 \text{ eV}}{1.602 \times 10^{-19} \text{ J}} \right) \\
 &= 54.0 \text{ eV}
 \end{aligned}$$

Discussion

This low energy means that these 0.167-nm electrons could be obtained by accelerating them through a 54.0-V electrostatic potential, an easy task. The results also confirm the assumption that the electrons are nonrelativistic, since their velocity is just over 1% of the speed of light and the kinetic energy is about 0.01% of the rest energy of an electron (0.511 MeV). If the electrons had turned out to be relativistic, we would have had to use more involved calculations employing relativistic formulas.

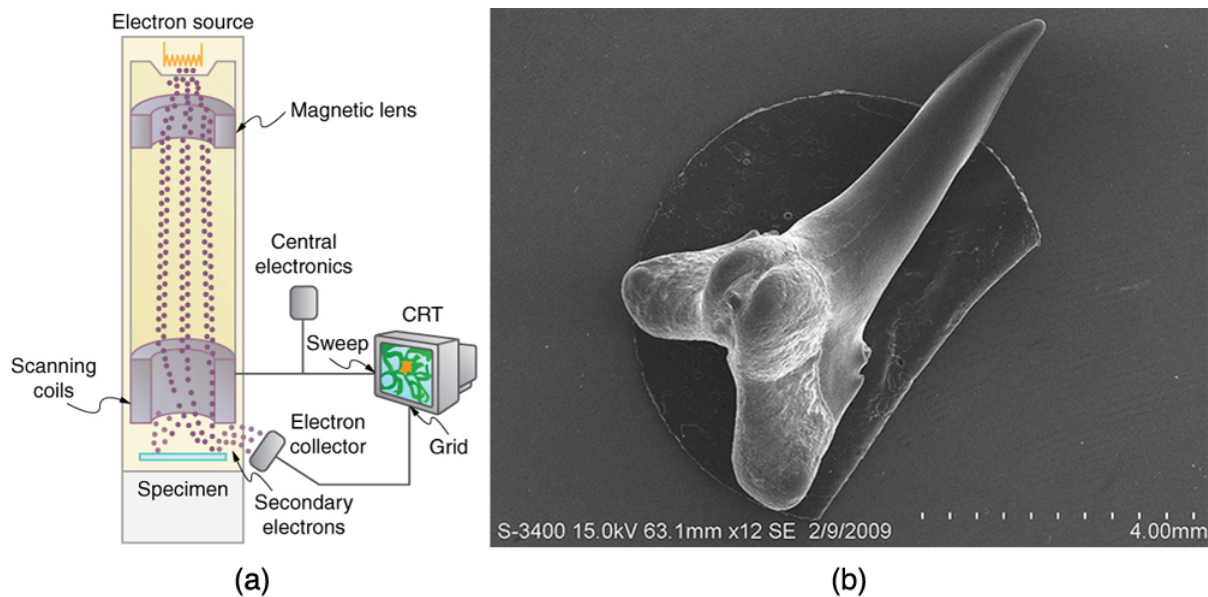
Electron Microscopes

One consequence or use of the wave nature of matter is found in the electron microscope. As we have discussed, there is a limit to the detail observed with any probe having a wavelength. Resolution, or observable detail, is limited to about one wavelength. Since a potential of only 54 V can produce electrons with sub-nanometer wavelengths, it is easy to get electrons with much smaller wavelengths than those of visible light (hundreds of nanometers). Electron microscopes can, thus, be constructed to detect much smaller details than optical microscopes. (See [\[link\]](#).)

There are basically two types of electron microscopes. The transmission electron microscope (TEM) accelerates electrons that are emitted from a hot filament (the cathode). The beam is broadened and then passes through the sample. A magnetic lens focuses the beam image onto a fluorescent screen, a photographic plate, or (most probably) a CCD (light sensitive camera), from which it is transferred to a computer. The TEM is similar to the optical microscope, but it requires a thin sample examined in a vacuum. However it can resolve details as small as 0.1 nm (10^{-10} m), providing magnifications

of 100 million times the size of the original object. The TEM has allowed us to see individual atoms and structure of cell nuclei.

The scanning electron microscope (SEM) provides images by using secondary electrons produced by the primary beam interacting with the surface of the sample (see [\[link\]](#)). The SEM also uses magnetic lenses to focus the beam onto the sample. However, it moves the beam around electrically to “scan” the sample in the x and y directions. A CCD detector is used to process the data for each electron position, producing images like the one at the beginning of this chapter. The SEM has the advantage of not requiring a thin sample and of providing a 3-D view. However, its resolution is about ten times less than a TEM.



Schematic of a scanning electron microscope (SEM) (a) used to observe small details, such as those seen in this image of a tooth of a *Himipristis*, a type of shark (b). (credit: Dallas Krentzel, Flickr)

Electrons were the first particles with mass to be directly confirmed to have the wavelength proposed by de Broglie. Subsequently, protons, helium nuclei, neutrons, and many others have been observed to exhibit

interference when they interact with objects having sizes similar to their de Broglie wavelength. The de Broglie wavelength for massless particles was well established in the 1920s for photons, and it has since been observed that all massless particles have a de Broglie wavelength $\lambda = h/p$. The wave nature of all particles is a universal characteristic of nature. We shall see in following sections that implications of the de Broglie wavelength include the quantization of energy in atoms and molecules, and an alteration of our basic view of nature on the microscopic scale. The next section, for example, shows that there are limits to the precision with which we may make predictions, regardless of how hard we try. There are even limits to the precision with which we may measure an object's location or energy.

Note:

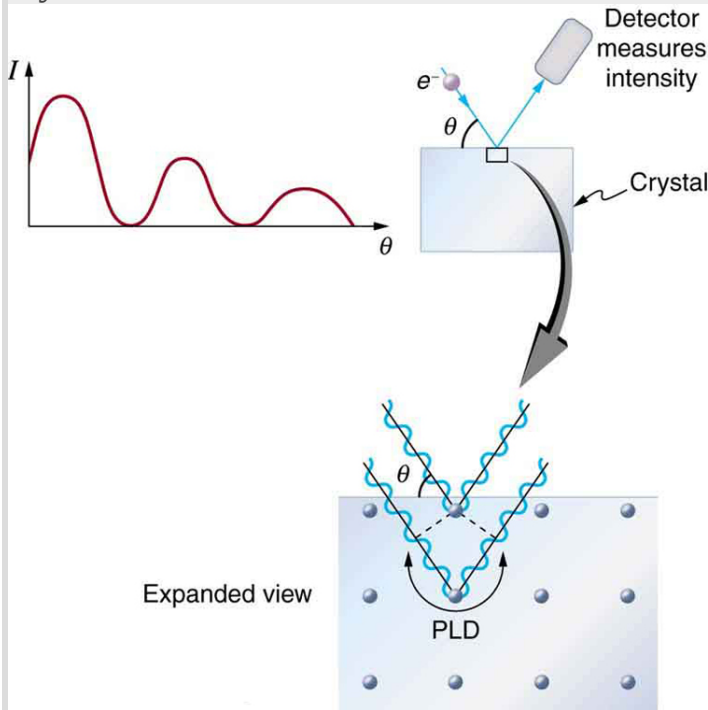
Making Connections: A Submicroscopic Diffraction Grating

The wave nature of matter allows it to exhibit all the characteristics of other, more familiar, waves. Diffraction gratings, for example, produce diffraction patterns for light that depend on grating spacing and the wavelength of the light. This effect, as with most wave phenomena, is most pronounced when the wave interacts with objects having a size similar to its wavelength. For gratings, this is the spacing between multiple slits.)

When electrons interact with a system having a spacing similar to the electron wavelength, they show the same types of interference patterns as light does for diffraction gratings, as shown at top left in [\[link\]](#).

Atoms are spaced at regular intervals in a crystal as parallel planes, as shown in the bottom part of [\[link\]](#). The spacings between these planes act like the openings in a diffraction grating. At certain incident angles, the paths of electrons scattering from successive planes differ by one wavelength and, thus, interfere constructively. At other angles, the path length differences are not an integral wavelength, and there is partial to total destructive interference. This type of scattering from a large crystal with well-defined lattice planes can produce dramatic interference patterns. It is called *Bragg reflection*, for the father-and-son team who first explored and analyzed it in some detail. The expanded view also shows the path-length differences and indicates how these depend on incident angle θ in a

manner similar to the diffraction patterns for x rays reflecting from a crystal.



The diffraction pattern at top left is produced by scattering electrons from a crystal and is graphed as a function of incident angle relative to the regular array of atoms in a crystal, as shown at bottom. Electrons scattering from the second layer of atoms travel farther than those scattered from the top layer. If the path length difference (PLD) is an integral wavelength, there is constructive interference.

Let us take the spacing between parallel planes of atoms in the crystal to be d . As mentioned, if the path length difference (PLD) for the electrons is a whole number of wavelengths, there will be constructive interference—that is, $\text{PLD} = n\lambda (n = 1, 2, 3, \dots)$. Because $AB = BC = d \sin \theta$, we have constructive interference when $n\lambda = 2d \sin \theta$. This relationship is

called the *Bragg equation* and applies not only to electrons but also to x rays.

The wavelength of matter is a submicroscopic characteristic that explains a macroscopic phenomenon such as Bragg reflection. Similarly, the wavelength of light is a submicroscopic characteristic that explains the macroscopic phenomenon of diffraction patterns.

Section Summary

- Particles of matter also have a wavelength, called the de Broglie wavelength, given by $\lambda = \frac{h}{p}$, where p is momentum.
- Matter is found to have the same *interference characteristics* as any other wave.

Conceptual Questions

Exercise:

Problem:

How does the interference of water waves differ from the interference of electrons? How are they analogous?

Exercise:

Problem: Describe one type of evidence for the wave nature of matter.

Exercise:

Problem:

Describe one type of evidence for the particle nature of EM radiation.

Problems & Exercises

Exercise:

Problem:

At what velocity will an electron have a wavelength of 1.00 m?

Solution:

$$7.28 \times 10^{-4} \text{ m}$$

Exercise:**Problem:**

What is the wavelength of an electron moving at 3.00% of the speed of light?

Exercise:**Problem:**

At what velocity does a proton have a 6.00-fm wavelength (about the size of a nucleus)? Assume the proton is nonrelativistic. (1 femtometer = 10^{-15} m.)

Solution:

$$6.62 \times 10^7 \text{ m/s}$$

Exercise:**Problem:**

What is the velocity of a 0.400-kg billiard ball if its wavelength is 7.50 cm (large enough for it to interfere with other billiard balls)?

Exercise:**Problem:**

Find the wavelength of a proton moving at 1.00% of the speed of light.

Solution:

$$1.32 \times 10^{-13} \text{ m}$$

Exercise:**Problem:**

Experiments are performed with ultracold neutrons having velocities as small as 1.00 m/s. (a) What is the wavelength of such a neutron? (b) What is its kinetic energy in eV?

Exercise:**Problem:**

(a) Find the velocity of a neutron that has a 6.00-fm wavelength (about the size of a nucleus). Assume the neutron is nonrelativistic. (b) What is the neutron's kinetic energy in MeV?

Solution:

(a) $6.62 \times 10^7 \text{ m/s}$

(b) 22.9 MeV

Exercise:**Problem:**

What is the wavelength of an electron accelerated through a 30.0-kV potential, as in a TV tube?

Exercise:**Problem:**

What is the kinetic energy of an electron in a TEM having a 0.0100-nm wavelength?

Solution:

Equation: 15.1 keV

Exercise:

Problem:

(a) Calculate the velocity of an electron that has a wavelength of $1.00\text{ }\mu\text{m}$. (b) Through what voltage must the electron be accelerated to have this velocity?

Exercise:**Problem:**

The velocity of a proton emerging from a Van de Graaff accelerator is 25.0% of the speed of light. (a) What is the proton's wavelength? (b) What is its kinetic energy, assuming it is nonrelativistic? (c) What was the equivalent voltage through which it was accelerated?

Solution:

(a) 5.29 fm

(b) $4.70 \times 10^{-12}\text{ J}$

(c) 29.4 MV

Exercise:**Problem:**

The kinetic energy of an electron accelerated in an x-ray tube is 100 keV . Assuming it is nonrelativistic, what is its wavelength?

Exercise:**Problem: Unreasonable Results**

(a) Assuming it is nonrelativistic, calculate the velocity of an electron with a 0.100-fm wavelength (small enough to detect details of a nucleus). (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $7.28 \times 10^{12} \text{ m/s}$

(b) This is thousands of times the speed of light (an impossibility).

(c) The assumption that the electron is non-relativistic is unreasonable at this wavelength.

Glossary

de Broglie wavelength

the wavelength possessed by a particle of matter, calculated by

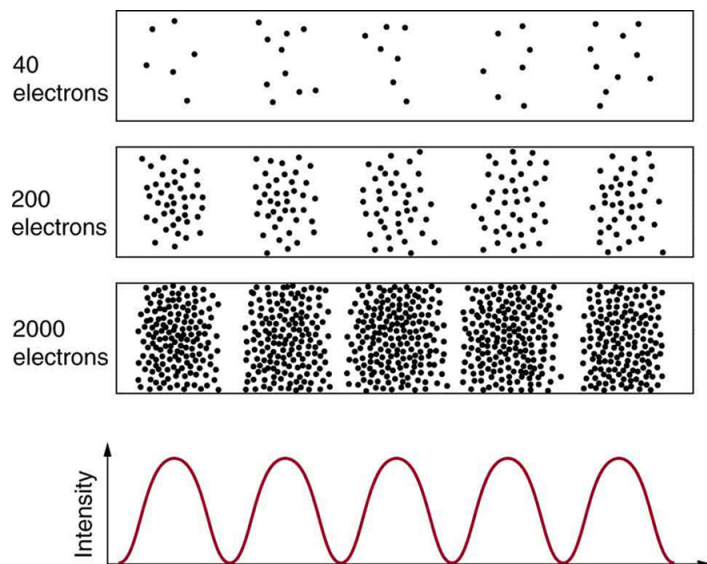
$$\lambda = h/p$$

Probability: The Heisenberg Uncertainty Principle

- Use both versions of Heisenberg's uncertainty principle in calculations.
- Explain the implications of Heisenberg's uncertainty principle for measurements.

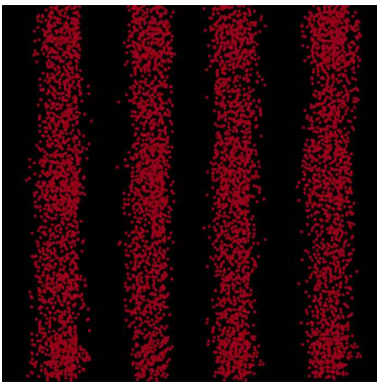
Probability Distribution

Matter and photons are waves, implying they are spread out over some distance. What is the position of a particle, such as an electron? Is it at the center of the wave? The answer lies in how you measure the position of an electron. Experiments show that you will find the electron at some definite location, unlike a wave. But if you set up exactly the same situation and measure it again, you will find the electron in a different location, often far outside any experimental uncertainty in your measurement. Repeated measurements will display a statistical distribution of locations that appears wavelike. (See [\[link\]](#).)

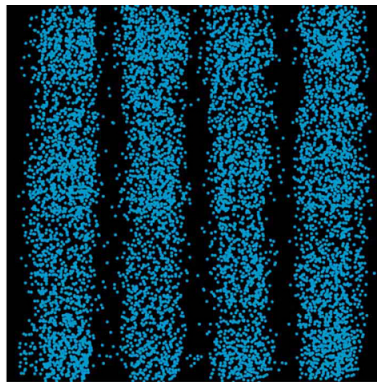


The building up of the diffraction pattern of electrons scattered from a crystal surface. Each electron arrives

at a definite location, which cannot be precisely predicted. The overall distribution shown at the bottom can be predicted as the diffraction of waves having the de Broglie wavelength of the electrons.



(a) Electrons



(b) Protons

Double-slit interference for electrons (a) and protons (b) is identical for equal wavelengths and equal slit separations. Both patterns are probability distributions in the sense that they are built up by individual particles traversing the apparatus, the paths of which are not individually predictable.

After de Broglie proposed the wave nature of matter, many physicists, including Schrödinger and Heisenberg, explored the consequences. The idea quickly emerged that, *because of its wave character, a particle's trajectory and destination cannot be precisely predicted for each particle individually*. However, each particle goes to a definite place (as illustrated in [\[link\]](#)). After compiling enough data, you get a distribution related to the

particle's wavelength and diffraction pattern. There is a certain *probability* of finding the particle at a given location, and the overall pattern is called a **probability distribution**. Those who developed quantum mechanics devised equations that predicted the probability distribution in various circumstances.

It is somewhat disquieting to think that you cannot predict exactly where an individual particle will go, or even follow it to its destination. Let us explore what happens if we try to follow a particle. Consider the double-slit patterns obtained for electrons and photons in [\[link\]](#). First, we note that these patterns are identical, following $d \sin \theta = m\lambda$, the equation for double-slit constructive interference developed in [Photon Energies and the Electromagnetic Spectrum](#), where d is the slit separation and λ is the electron or photon wavelength.

Both patterns build up statistically as individual particles fall on the detector. This can be observed for photons or electrons—for now, let us concentrate on electrons. You might imagine that the electrons are interfering with one another as any waves do. To test this, you can lower the intensity until there is never more than one electron between the slits and the screen. The same interference pattern builds up! This implies that a particle's probability distribution spans both slits, and the particles actually interfere with themselves. Does this also mean that the electron goes through both slits? An electron is a basic unit of matter that is not divisible. But it is a fair question, and so we should look to see if the electron traverses one slit or the other, or both. One possibility is to have coils around the slits that detect charges moving through them. What is observed is that an electron always goes through one slit or the other; it does not split to go through both. But there is a catch. If you determine that the electron went through one of the slits, you no longer get a double slit pattern—instead, you get single slit interference. There is no escape by using another method of determining which slit the electron went through. Knowing the particle went through one slit forces a single-slit pattern. If you do not observe which slit the electron goes through, you obtain a double-slit pattern.

Heisenberg Uncertainty

How does knowing which slit the electron passed through change the pattern? The answer is fundamentally important—*measurement affects the system being observed*. Information can be lost, and in some cases it is impossible to measure two physical quantities simultaneously to exact precision. For example, you can measure the position of a moving electron by scattering light or other electrons from it. Those probes have momentum themselves, and by scattering from the electron, they change its momentum *in a manner that loses information*. There is a limit to absolute knowledge, even in principle.



Werner Heisenberg was one of the best of those physicists who developed early quantum mechanics. Not only did his work enable a description of nature on the very small scale, it also changed our

view of the
availability of
knowledge.
Although he is
universally
recognized for his
brilliance and the
importance of his
work (he received
the Nobel Prize in
1932, for example),
Heisenberg
remained in
Germany during
World War II and
headed the German
effort to build a
nuclear bomb,
permanently
alienating himself
from most of the
scientific
community. (credit:
Author Unknown,
via Wikimedia
Commons)

It was Werner Heisenberg who first stated this limit to knowledge in 1929 as a result of his work on quantum mechanics and the wave characteristics of all particles. (See [\[link\]](#)). Specifically, consider simultaneously measuring the position and momentum of an electron (it could be any particle). There is an **uncertainty in position** Δx that is approximately equal to the wavelength of the particle. That is,
Equation:

$$\Delta x \approx \lambda.$$

As discussed above, a wave is not located at one point in space. If the electron's position is measured repeatedly, a spread in locations will be observed, implying an uncertainty in position Δx . To detect the position of the particle, we must interact with it, such as having it collide with a detector. In the collision, the particle will lose momentum. This change in momentum could be anywhere from close to zero to the total momentum of the particle, $p = h/\lambda$. It is not possible to tell how much momentum will be transferred to a detector, and so there is an **uncertainty in momentum** Δp , too. In fact, the uncertainty in momentum may be as large as the momentum itself, which in equation form means that

Equation:

$$\Delta p \approx \frac{h}{\lambda}.$$

The uncertainty in position can be reduced by using a shorter-wavelength electron, since $\Delta x \approx \lambda$. But shortening the wavelength increases the uncertainty in momentum, since $\Delta p \approx h/\lambda$. Conversely, the uncertainty in momentum can be reduced by using a longer-wavelength electron, but this increases the uncertainty in position. Mathematically, you can express this trade-off by multiplying the uncertainties. The wavelength cancels, leaving

Equation:

$$\Delta x \Delta p \approx h.$$

So if one uncertainty is reduced, the other must increase so that their product is $\approx h$.

With the use of advanced mathematics, Heisenberg showed that the best that can be done in a *simultaneous measurement of position and momentum* is

Equation:

$$\Delta x \Delta p \geq \frac{h}{4\pi}.$$

This is known as the **Heisenberg uncertainty principle**. It is impossible to measure position x and momentum p simultaneously with uncertainties Δx and Δp that multiply to be less than $h/4\pi$. Neither uncertainty can be zero. Neither uncertainty can become small without the other becoming large. A small wavelength allows accurate position measurement, but it increases the momentum of the probe to the point that it further disturbs the momentum of a system being measured. For example, if an electron is scattered from an atom and has a wavelength small enough to detect the position of electrons in the atom, its momentum can knock the electrons from their orbits in a manner that loses information about their original motion. It is therefore impossible to follow an electron in its orbit around an atom. If you measure the electron's position, you will find it in a definite location, but the atom will be disrupted. Repeated measurements on identical atoms will produce interesting probability distributions for electrons around the atom, but they will not produce motion information. The probability distributions are referred to as electron clouds or orbitals. The shapes of these orbitals are often shown in general chemistry texts and are discussed in [The Wave Nature of Matter Causes Quantization](#).

Example:

Heisenberg Uncertainty Principle in Position and Momentum for an Atom

(a) If the position of an electron in an atom is measured to an accuracy of 0.0100 nm, what is the electron's uncertainty in velocity? (b) If the electron has this velocity, what is its kinetic energy in eV?

Strategy

The uncertainty in position is the accuracy of the measurement, or $\Delta x = 0.0100$ nm. Thus the smallest uncertainty in momentum Δp can be calculated using $\Delta x \Delta p \geq h/4\pi$. Once the uncertainty in momentum Δp is found, the uncertainty in velocity can be found from $\Delta p = m\Delta v$.

Solution for (a)

Using the equals sign in the uncertainty principle to express the minimum uncertainty, we have

Equation:

$$\Delta x \Delta p = \frac{h}{4\pi}.$$

Solving for Δp and substituting known values gives

Equation:

$$\Delta p = \frac{h}{4\pi\Delta x} = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{4\pi(1.00 \times 10^{-11} \text{ m})} = 5.28 \times 10^{-24} \text{ kg} \cdot \text{m/s}.$$

Thus,

Equation:

$$\Delta p = 5.28 \times 10^{-24} \text{ kg} \cdot \text{m/s} = m\Delta v.$$

Solving for Δv and substituting the mass of an electron gives

Equation:

$$\Delta v = \frac{\Delta p}{m} = \frac{5.28 \times 10^{-24} \text{ kg} \cdot \text{m/s}}{9.11 \times 10^{-31} \text{ kg}} = 5.79 \times 10^6 \text{ m/s}.$$

Solution for (b)

Although large, this velocity is not highly relativistic, and so the electron's kinetic energy is

Equation:

$$\begin{aligned} \text{KE}_e &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(5.79 \times 10^6 \text{ m/s})^2 \\ &= (1.53 \times 10^{-17} \text{ J})\left(\frac{1 \text{ eV}}{1.60 \times 10^{-19} \text{ J}}\right) = 95.5 \text{ eV}. \end{aligned}$$

Discussion

Since atoms are roughly 0.1 nm in size, knowing the position of an electron to 0.0100 nm localizes it reasonably well inside the atom. This

would be like being able to see details one-tenth the size of the atom. But the consequent uncertainty in velocity is large. You certainly could not follow it very well if its velocity is so uncertain. To get a further idea of how large the uncertainty in velocity is, we assumed the velocity of the electron was equal to its uncertainty and found this gave a kinetic energy of 95.5 eV. This is significantly greater than the typical energy difference between levels in atoms (see [\[link\]](#)), so that it is impossible to get a meaningful energy for the electron if we know its position even moderately well.

Why don't we notice Heisenberg's uncertainty principle in everyday life? The answer is that Planck's constant is very small. Thus the lower limit in the uncertainty of measuring the position and momentum of large objects is negligible. We can detect sunlight reflected from Jupiter and follow the planet in its orbit around the Sun. The reflected sunlight alters the momentum of Jupiter and creates an uncertainty in its momentum, but this is totally negligible compared with Jupiter's huge momentum. The correspondence principle tells us that the predictions of quantum mechanics become indistinguishable from classical physics for large objects, which is the case here.

Heisenberg Uncertainty for Energy and Time

There is another form of **Heisenberg's uncertainty principle** for *simultaneous measurements of energy and time*. In equation form,

Equation:

$$\Delta E \Delta t \geq \frac{h}{4\pi},$$

where ΔE is the **uncertainty in energy** and Δt is the **uncertainty in time**. This means that within a time interval Δt , it is not possible to measure energy precisely—there will be an uncertainty ΔE in the measurement. In order to measure energy more precisely (to make ΔE smaller), we must

increase Δt . This time interval may be the amount of time we take to make the measurement, or it could be the amount of time a particular state exists, as in the next [\[link\]](#).

Example:**Heisenberg Uncertainty Principle for Energy and Time for an Atom**

An atom in an excited state temporarily stores energy. If the lifetime of this excited state is measured to be 1.0×10^{-10} s, what is the minimum uncertainty in the energy of the state in eV?

Strategy

The minimum uncertainty in energy ΔE is found by using the equals sign in $\Delta E \Delta t \geq h/4\pi$ and corresponds to a reasonable choice for the uncertainty in time. The largest the uncertainty in time can be is the full lifetime of the excited state, or $\Delta t = 1.0 \times 10^{-10}$ s.

Solution

Solving the uncertainty principle for ΔE and substituting known values gives

Equation:

$$\Delta E = \frac{h}{4\pi\Delta t} = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{4\pi(1.0 \times 10^{-10} \text{ s})} = 5.3 \times 10^{-25} \text{ J}.$$

Now converting to eV yields

Equation:

$$\Delta E = (5.3 \times 10^{-25} \text{ J}) \left(\frac{1 \text{ eV}}{1.6 \times 10^{-19} \text{ J}} \right) = 3.3 \times 10^{-6} \text{ eV}.$$

Discussion

The lifetime of 10^{-10} s is typical of excited states in atoms—on human time scales, they quickly emit their stored energy. An uncertainty in energy of only a few millionths of an eV results. This uncertainty is small compared with typical excitation energies in atoms, which are on the order of 1 eV. So here the uncertainty principle limits the accuracy with which

we can measure the lifetime and energy of such states, but not very significantly.

The uncertainty principle for energy and time can be of great significance if the lifetime of a system is very short. Then Δt is very small, and ΔE is consequently very large. Some nuclei and exotic particles have extremely short lifetimes (as small as 10^{-25} s), causing uncertainties in energy as great as many GeV (10^9 eV). Stored energy appears as increased rest mass, and so this means that there is significant uncertainty in the rest mass of short-lived particles. When measured repeatedly, a spread of masses or decay energies are obtained. The spread is ΔE . You might ask whether this uncertainty in energy could be avoided by not measuring the lifetime. The answer is no. Nature knows the lifetime, and so its brevity affects the energy of the particle. This is so well established experimentally that the uncertainty in decay energy is used to calculate the lifetime of short-lived states. Some nuclei and particles are so short-lived that it is difficult to measure their lifetime. But if their decay energy can be measured, its spread is ΔE , and this is used in the uncertainty principle ($\Delta E \Delta t \geq h/4\pi$) to calculate the lifetime Δt .

There is another consequence of the uncertainty principle for energy and time. If energy is uncertain by ΔE , then conservation of energy can be violated by ΔE for a time Δt . Neither the physicist nor nature can tell that conservation of energy has been violated, if the violation is temporary and smaller than the uncertainty in energy. While this sounds innocuous enough, we shall see in later chapters that it allows the temporary creation of matter from nothing and has implications for how nature transmits forces over very small distances.

Finally, note that in the discussion of particles and waves, we have stated that individual measurements produce precise or particle-like results. A definite position is determined each time we observe an electron, for example. But repeated measurements produce a spread in values consistent with wave characteristics. The great theoretical physicist Richard Feynman (1918–1988) commented, “What there are, are particles.” When you

observe enough of them, they distribute themselves as you would expect for a wave phenomenon. However, what there are as they travel we cannot tell because, when we do try to measure, we affect the traveling.

Section Summary

- Matter is found to have the same interference characteristics as any other wave.
- There is now a probability distribution for the location of a particle rather than a definite position.
- Another consequence of the wave character of all particles is the Heisenberg uncertainty principle, which limits the precision with which certain physical quantities can be known simultaneously. For position and momentum, the uncertainty principle is $\Delta x \Delta p \geq \frac{h}{4\pi}$, where Δx is the uncertainty in position and Δp is the uncertainty in momentum.
- For energy and time, the uncertainty principle is $\Delta E \Delta t \geq \frac{h}{4\pi}$ where ΔE is the uncertainty in energy and Δt is the uncertainty in time.
- These small limits are fundamentally important on the quantum-mechanical scale.

Conceptual Questions

Exercise:

Problem:

What is the Heisenberg uncertainty principle? Does it place limits on what can be known?

Problems & Exercises

Exercise:

Problem:

(a) If the position of an electron in a membrane is measured to an accuracy of $1.00\ \mu\text{m}$, what is the electron's minimum uncertainty in velocity? (b) If the electron has this velocity, what is its kinetic energy in eV? (c) What are the implications of this energy, comparing it to typical molecular binding energies?

Solution:

(a) $57.9\ \text{m/s}$

(b) $9.55 \times 10^{-9}\ \text{eV}$

(c) From [\[link\]](#), we see that typical molecular binding energies range from about 1 eV to 10 eV, therefore the result in part (b) is approximately 9 orders of magnitude smaller than typical molecular binding energies.

Exercise:**Problem:**

(a) If the position of a chlorine ion in a membrane is measured to an accuracy of $1.00\ \mu\text{m}$, what is its minimum uncertainty in velocity, given its mass is $5.86 \times 10^{-26}\ \text{kg}$? (b) If the ion has this velocity, what is its kinetic energy in eV, and how does this compare with typical molecular binding energies?

Exercise:**Problem:**

Suppose the velocity of an electron in an atom is known to an accuracy of $2.0 \times 10^3\ \text{m/s}$ (reasonably accurate compared with orbital velocities). What is the electron's minimum uncertainty in position, and how does this compare with the approximate 0.1-nm size of the atom?

Solution:

29 nm,

290 times greater

Exercise:**Problem:**

The velocity of a proton in an accelerator is known to an accuracy of 0.250% of the speed of light. (This could be small compared with its velocity.) What is the smallest possible uncertainty in its position?

Exercise:**Problem:**

A relatively long-lived excited state of an atom has a lifetime of 3.00 ms. What is the minimum uncertainty in its energy?

Solution:

$$1.10 \times 10^{-13} \text{ eV}$$

Exercise:**Problem:**

(a) The lifetime of a highly unstable nucleus is 10^{-20} s. What is the smallest uncertainty in its decay energy? (b) Compare this with the rest energy of an electron.

Exercise:**Problem:**

The decay energy of a short-lived particle has an uncertainty of 1.0 MeV due to its short lifetime. What is the smallest lifetime it can have?

Solution:

$$3.3 \times 10^{-22} \text{ s}$$

Exercise:**Problem:**

The decay energy of a short-lived nuclear excited state has an uncertainty of 2.0 eV due to its short lifetime. What is the smallest lifetime it can have?

Exercise:**Problem:**

What is the approximate uncertainty in the mass of a muon, as determined from its decay lifetime?

Solution:

$$2.66 \times 10^{-46} \text{ kg}$$

Exercise:**Problem:**

Derive the approximate form of Heisenberg's uncertainty principle for energy and time, $\Delta E \Delta t \approx h$, using the following arguments: Since the position of a particle is uncertain by $\Delta x \approx \lambda$, where λ is the wavelength of the photon used to examine it, there is an uncertainty in the time the photon takes to traverse Δx . Furthermore, the photon has an energy related to its wavelength, and it can transfer some or all of this energy to the object being examined. Thus the uncertainty in the energy of the object is also related to λ . Find Δt and ΔE ; then multiply them to give the approximate uncertainty principle.

Glossary**Heisenberg's uncertainty principle**

a fundamental limit to the precision with which pairs of quantities (momentum and position, and energy and time) can be measured

uncertainty in energy

lack of precision or lack of knowledge of precise results in measurements of energy

uncertainty in time

lack of precision or lack of knowledge of precise results in measurements of time

uncertainty in momentum

lack of precision or lack of knowledge of precise results in measurements of momentum

uncertainty in position

lack of precision or lack of knowledge of precise results in measurements of position

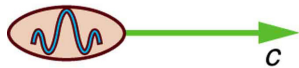
probability distribution

the overall spatial distribution of probabilities to find a particle at a given location

The Particle-Wave Duality Reviewed

- Explain the concept of particle-wave duality, and its scope.

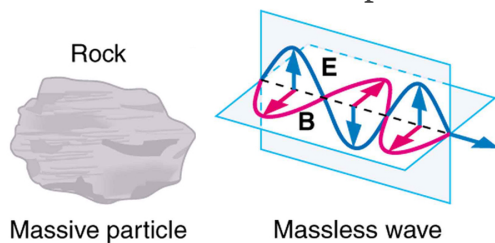
Particle-wave duality—the fact that all particles have wave properties—is one of the cornerstones of quantum mechanics. We first came across it in the treatment of photons, those particles of EM radiation that exhibit both particle and wave properties, but not at the same time. Later it was noted that particles of matter have wave properties as well. The dual properties of particles and waves are found for all particles, whether massless like photons, or having a mass like electrons. (See [\[link\]](#).)



On a quantum-mechanical scale (i.e., very small), particles with and without mass have wave properties. For example, both electrons and photons have wavelengths but also behave as particles.

There are many submicroscopic particles in nature. Most have mass and are expected to act as particles, or the smallest units of matter. All these masses have wave properties, with wavelengths given by the de Broglie relationship $\lambda = h/p$. So, too, do combinations of these particles, such as nuclei, atoms, and molecules. As a combination of masses becomes large, particularly if it is large enough to be called macroscopic, its wave nature becomes difficult to observe. This is consistent with our common experience with matter.

Some particles in nature are massless. We have only treated the photon so far, but all massless entities travel at the speed of light, have a wavelength, and exhibit particle and wave behaviors. They have momentum given by a rearrangement of the de Broglie relationship, $p = h/\lambda$. In large combinations of these massless particles (such large combinations are common only for photons or EM waves), there is mostly wave behavior upon detection, and the particle nature becomes difficult to observe. This is also consistent with experience. (See [\[link\]](#).)



On a classical scale (macroscopic), particles with mass behave as particles and not as waves. Particles without mass act as waves and not as particles.

The particle-wave duality is a universal attribute. It is another connection between matter and energy. Not only has modern physics been able to

describe nature for high speeds and small sizes, it has also discovered new connections and symmetries. There is greater unity and symmetry in nature than was known in the classical era—but they were dreamt of. A beautiful poem written by the English poet William Blake some two centuries ago contains the following four lines:

To see the World in a Grain of Sand

And a Heaven in a Wild Flower

Hold Infinity in the palm of your hand

And Eternity in an hour

Integrated Concepts

The problem set for this section involves concepts from this chapter and several others. Physics is most interesting when applied to general situations involving more than a narrow set of physical principles. For example, photons have momentum, hence the relevance of [Linear Momentum and Collisions](#). The following topics are involved in some or all of the problems in this section:

- [Dynamics: Newton's Laws of Motion](#)
- [Work, Energy, and Energy Resources](#)
- [Linear Momentum and Collisions](#)
- [Heat and Heat Transfer Methods](#)
- [Electric Potential and Electric Field](#)
- [Electric Current, Resistance, and Ohm's Law](#)
- [Wave Optics](#)
- [Special Relativity](#)

Note:

Problem-Solving Strategy

1. Identify which physical principles are involved.

2. Solve the problem using strategies outlined in the text.

[\[link\]](#) illustrates how these strategies are applied to an integrated-concept problem.

Example:

Recoil of a Dust Particle after Absorbing a Photon

The following topics are involved in this integrated concepts worked example:

Photons (quantum mechanics)

Linear Momentum

Topics

A 550-nm photon (visible light) is absorbed by a 1.00- μg particle of dust in outer space. (a) Find the momentum of such a photon. (b) What is the recoil velocity of the particle of dust, assuming it is initially at rest?

Strategy Step 1

To solve an *integrated-concept problem*, such as those following this example, we must first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example asks for the *momentum of a photon*, a topic of the present chapter. Part (b) considers *recoil following a collision*, a topic of [Linear Momentum and Collisions](#).

Strategy Step 2

The following solutions to each part of the example illustrate how specific problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so on.

Solution for (a)

The momentum of a photon is related to its wavelength by the equation:

Equation:

$$p = \frac{h}{\lambda}.$$

Entering the known value for Planck's constant h and given the wavelength λ , we obtain

Equation:

$$\begin{aligned} p &= \frac{6.63 \times 10^{-34} \text{ J}\cdot\text{s}}{550 \times 10^{-9} \text{ m}} \\ &= 1.21 \times 10^{-27} \text{ kg} \cdot \text{m/s}. \end{aligned}$$

Discussion for (a)

This momentum is small, as expected from discussions in the text and the fact that photons of visible light carry small amounts of energy and momentum compared with those carried by macroscopic objects.

Solution for (b)

Conservation of momentum in the absorption of this photon by a grain of dust can be analyzed using the equation:

Equation:

$$p_1 + p_2 = p'_1 + p'_2 (F_{\text{net}} = 0).$$

The net external force is zero, since the dust is in outer space. Let 1 represent the photon and 2 the dust particle. Before the collision, the dust is at rest (relative to some observer); after the collision, there is no photon (it is absorbed). So conservation of momentum can be written

Equation:

$$p_1 = p'_2 = mv,$$

where p_1 is the photon momentum before the collision and p'_2 is the dust momentum after the collision. The mass and recoil velocity of the dust are

m and v , respectively. Solving this for v , the requested quantity, yields

Equation:

$$v = \frac{p}{m},$$

where p is the photon momentum found in part (a). Entering known values (noting that a microgram is 10^{-9} kg) gives

Equation:

$$\begin{aligned} v &= \frac{1.21 \times 10^{-27} \text{ kg}\cdot\text{m/s}}{1.00 \times 10^{-9} \text{ kg}} \\ &= 1.21 \times 10^{-18} \text{ m/s.} \end{aligned}$$

Discussion

The recoil velocity of the particle of dust is extremely small. As we have noted, however, there are immense numbers of photons in sunlight and other macroscopic sources. In time, collisions and absorption of many photons could cause a significant recoil of the dust, as observed in comet tails.

Section Summary

- The particle-wave duality refers to the fact that all particles—those with mass and those without mass—have wave characteristics.
- This is a further connection between mass and energy.

Conceptual Questions

Exercise:

Problem:

In what ways are matter and energy related that were not known before the development of relativity and quantum mechanics?

Problems & Exercises

Exercise:

Problem: Integrated Concepts

The 54.0-eV electron in [\[link\]](#) has a 0.167-nm wavelength. If such electrons are passed through a double slit and have their first maximum at an angle of 25.0° , what is the slit separation d ?

Solution:

0.395 nm

Exercise:

Problem: Integrated Concepts

An electron microscope produces electrons with a 2.00-pm wavelength. If these are passed through a 1.00-nm single slit, at what angle will the first diffraction minimum be found?

Exercise:

Problem: Integrated Concepts

A certain heat lamp emits 200 W of mostly IR radiation averaging 1500 nm in wavelength. (a) What is the average photon energy in joules? (b) How many of these photons are required to increase the temperature of a person's shoulder by 2.0°C , assuming the affected mass is 4.0 kg with a specific heat of $0.83 \text{ kcal/kg}\cdot^\circ\text{C}$. Also assume no other significant heat transfer. (c) How long does this take?

Solution:

(a) $1.3 \times 10^{-19} \text{ J}$

(b) 2.1×10^{23}

(c) $1.4 \times 10^2 \text{ s}$

Exercise:

Problem: Integrated Concepts

On its high power setting, a microwave oven produces 900 W of 2560 MHz microwaves. (a) How many photons per second is this? (b) How many photons are required to increase the temperature of a 0.500-kg mass of pasta by 45.0°C , assuming a specific heat of $0.900 \text{ kcal/kg} \cdot ^\circ\text{C}$? Neglect all other heat transfer. (c) How long must the microwave operator wait for their pasta to be ready?

Exercise:

Problem: Integrated Concepts

(a) Calculate the amount of microwave energy in joules needed to raise the temperature of 1.00 kg of soup from 20.0°C to 100°C . (b) What is the total momentum of all the microwave photons it takes to do this? (c) Calculate the velocity of a 1.00-kg mass with the same momentum. (d) What is the kinetic energy of this mass?

Solution:

(a) $3.35 \times 10^5 \text{ J}$

(b) $1.12 \times 10^{-3} \text{ kg} \cdot \text{m/s}$

(c) $1.12 \times 10^{-3} \text{ m/s}$

(d) $6.23 \times 10^{-7} \text{ J}$

Exercise:

Problem: Integrated Concepts

- (a) What is γ for an electron emerging from the Stanford Linear Accelerator with a total energy of 50.0 GeV? (b) Find its momentum. (c) What is the electron's wavelength?

Exercise:

Problem: Integrated Concepts

- (a) What is γ for a proton having an energy of 1.00 TeV, produced by the Fermilab accelerator? (b) Find its momentum. (c) What is the proton's wavelength?

Solution:

- (a) 1.06×10^3
(b) $5.33 \times 10^{-16} \text{ kg} \cdot \text{m/s}$
(c) $1.24 \times 10^{-18} \text{ m}$

Exercise:

Problem: Integrated Concepts

An electron microscope passes 1.00-pm-wavelength electrons through a circular aperture 2.00 μm in diameter. What is the angle between two just-resolvable point sources for this microscope?

Exercise:

Problem: Integrated Concepts

- (a) Calculate the velocity of electrons that form the same pattern as 450-nm light when passed through a double slit. (b) Calculate the kinetic energy of each and compare them. (c) Would either be easier to generate than the other? Explain.

Solution:

(a) $1.62 \times 10^3 \text{ m/s}$

(b) $4.42 \times 10^{-19} \text{ J}$ for photon, $1.19 \times 10^{-24} \text{ J}$ for electron, photon energy is 3.71×10^5 times greater

(c) The light is easier to make because 450-nm light is blue light and therefore easy to make. Creating electrons with $7.43 \text{ } \mu\text{eV}$ of energy would not be difficult, but would require a vacuum.

Exercise:

Problem: Integrated Concepts

(a) What is the separation between double slits that produces a second-order minimum at 45.0° for 650-nm light? (b) What slit separation is needed to produce the same pattern for 1.00-keV protons.

Solution:

(a) $2.30 \times 10^{-6} \text{ m}$

(b) $3.20 \times 10^{-12} \text{ m}$

Exercise:

Problem: Integrated Concepts

A laser with a power output of 2.00 mW at a wavelength of 400 nm is projected onto calcium metal. (a) How many electrons per second are ejected? (b) What power is carried away by the electrons, given that the binding energy is 2.71 eV? (c) Calculate the current of ejected electrons. (d) If the photoelectric material is electrically insulated and acts like a 2.00-pF capacitor, how long will current flow before the capacitor voltage stops it?

Exercise:

Problem: Integrated Concepts

One problem with x rays is that they are not sensed. Calculate the temperature increase of a researcher exposed in a few seconds to a nearly fatal accidental dose of x rays under the following conditions. The energy of the x-ray photons is 200 keV, and 4.00×10^{13} of them are absorbed per kilogram of tissue, the specific heat of which is $0.830 \text{ kcal/kg} \cdot ^\circ\text{C}$. (Note that medical diagnostic x-ray machines *cannot* produce an intensity this great.)

Solution:

$$3.69 \times 10^{-4} \text{ }^\circ\text{C}$$

Exercise:

Problem: Integrated Concepts

A 1.00-fm photon has a wavelength short enough to detect some information about nuclei. (a) What is the photon momentum? (b) What is its energy in joules and MeV? (c) What is the (relativistic) velocity of an electron with the same momentum? (d) Calculate the electron's kinetic energy.

Exercise:

Problem: Integrated Concepts

The momentum of light is exactly reversed when reflected straight back from a mirror, assuming negligible recoil of the mirror. Thus the change in momentum is twice the photon momentum. Suppose light of intensity 1.00 kW/m^2 reflects from a mirror of area 2.00 m^2 . (a) Calculate the energy reflected in 1.00 s. (b) What is the momentum imparted to the mirror? (c) Using the most general form of Newton's second law, what is the force on the mirror? (d) Does the assumption of no mirror recoil seem reasonable?

Solution:

(a) 2.00 kJ

(b) $1.33 \times 10^{-5} \text{ kg} \cdot \text{m/s}$

(c) $1.33 \times 10^{-5} \text{ N}$

(d) yes

Exercise:

Problem: Integrated Concepts

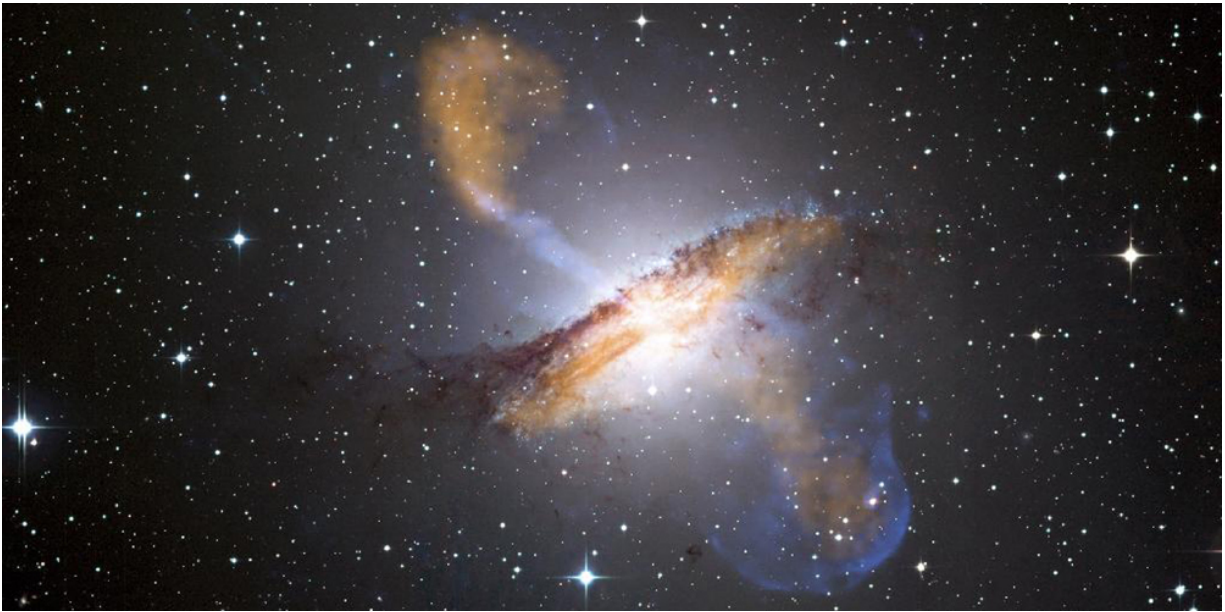
Sunlight above the Earth's atmosphere has an intensity of 1.30 kW/m^2 . If this is reflected straight back from a mirror that has only a small recoil, the light's momentum is exactly reversed, giving the mirror twice the incident momentum. (a) Calculate the force per square meter of mirror. (b) Very low mass mirrors can be constructed in the near weightlessness of space, and attached to a spaceship to sail it. Once done, the average mass per square meter of the spaceship is 0.100 kg . Find the acceleration of the spaceship if all other forces are balanced. (c) How fast is it moving 24 hours later?

Introduction to Frontiers of Physics

class="introduction"

This galaxy is
ejecting huge jets of
matter, powered by
an immensely
massive black hole
at its center. (credit:

X-ray:
NASA/CXC/CfA/R.
Kraft et al.)



Frontiers are exciting. There is mystery, surprise, adventure, and discovery. The satisfaction of finding the answer to a question is made keener by the fact that the answer always leads to a new question. The picture of nature becomes more complete, yet nature retains its sense of mystery and never loses its ability to awe us. The view of physics is beautiful looking both backward and forward in time. What marvelous patterns we have discovered. How clever nature seems in its rules and connections. How awesome. And we continue looking ever deeper and ever further, probing

the basic structure of matter, energy, space, and time and wondering about the scope of the universe, its beginnings and future.

You are now in a wonderful position to explore the forefronts of physics, both the new discoveries and the unanswered questions. With the concepts, qualitative and quantitative, the problem-solving skills, the feeling for connections among topics, and all the rest you have mastered, you can more deeply appreciate and enjoy the brief treatments that follow. Years from now you will still enjoy the quest with an insight all the greater for your efforts.

Cosmology and Particle Physics

- Discuss the expansion of the universe.
- Explain the Big Bang.

Look at the sky on some clear night when you are away from city lights. There you will see thousands of individual stars and a faint glowing background of millions more. The Milky Way, as it has been called since ancient times, is an arm of our galaxy of stars—the word *galaxy* coming from the Greek word *galaxias*, meaning milky. We know a great deal about our Milky Way galaxy and of the billions of other galaxies beyond its fringes. But they still provoke wonder and awe (see [\[link\]](#)). And there are still many questions to be answered. Most remarkable when we view the universe on the large scale is that once again explanations of its character and evolution are tied to the very small scale. Particle physics and the questions being asked about the very small scales may also have their answers in the very large scales.

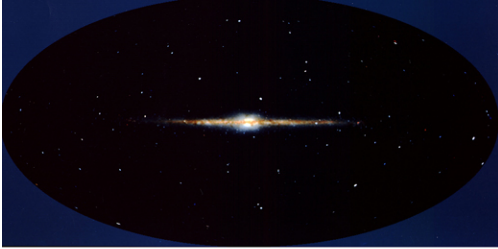


Take a moment to contemplate these clusters of galaxies, photographed by the Hubble Space Telescope. Trillions of stars linked by gravity in fantastic forms, glowing with light and showing evidence of undiscovered matter. What are they like, these myriad stars? How did they evolve? What can

they tell us of matter, energy,
space, and time? (credit: NASA,
ESA, K. Sharon (Tel Aviv
University) and E. Ofek
(Caltech))

As has been noted in numerous Things Great and Small vignettes, this is not the first time the large has been explained by the small and vice versa. Newton realized that the nature of gravity on Earth that pulls an apple to the ground could explain the motion of the moon and planets so much farther away. Minute atoms and molecules explain the chemistry of substances on a much larger scale. Decays of tiny nuclei explain the hot interior of the Earth. Fusion of nuclei likewise explains the energy of stars. Today, the patterns in particle physics seem to be explaining the evolution and character of the universe. And the nature of the universe has implications for unexplored regions of particle physics.

Cosmology is the study of the character and evolution of the universe. What are the major characteristics of the universe as we know them today? First, there are approximately 10^{11} galaxies in the observable part of the universe. An average galaxy contains more than 10^{11} stars, with our Milky Way galaxy being larger than average, both in its number of stars and its dimensions. Ours is a spiral-shaped galaxy with a diameter of about 100,000 light years and a thickness of about 2000 light years in the arms with a central bulge about 10,000 light years across. The Sun lies about 30,000 light years from the center near the galactic plane. There are significant clouds of gas, and there is a halo of less-dense regions of stars surrounding the main body. (See [\[link\]](#).) Evidence strongly suggests the existence of a large amount of additional matter in galaxies that does not produce light—the mysterious dark matter we shall later discuss.



(a)



(b)



(c)

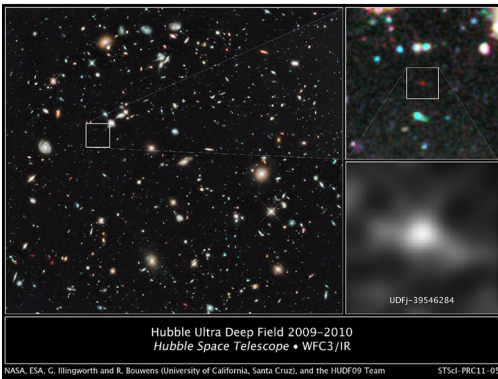
The Milky Way galaxy is typical of large spiral galaxies in its size, its shape, and the presence of gas and dust. We are fortunate to be in a location where we can see out of the galaxy and observe the vastly larger and fascinating universe

around us. (a) Side view.
(b) View from above. (c)
The Milky Way as seen
from Earth. (credits: (a)
NASA, (b) Nick Risinger,
(c) Andy)

Distances are great even within our galaxy and are measured in light years (the distance traveled by light in one year). The average distance between galaxies is on the order of a million light years, but it varies greatly with galaxies forming clusters such as shown in [\[link\]](#). The Magellanic Clouds, for example, are small galaxies close to our own, some 160,000 light years from Earth. The Andromeda galaxy is a large spiral galaxy like ours and lies 2 million light years away. It is just visible to the naked eye as an extended glow in the Andromeda constellation. Andromeda is the closest large galaxy in our local group, and we can see some individual stars in it with our larger telescopes. The most distant known galaxy is 14 billion light years from Earth—a truly incredible distance. (See [\[link\]](#).)



(a)



(b)

(a) Andromeda is the closest large galaxy, at 2 million light years distance, and is very similar to our Milky Way. The blue regions harbor young and emerging stars, while dark streaks are vast clouds of gas and dust. A smaller satellite galaxy is clearly visible.

(b) The box indicates what may be the most distant known galaxy, estimated to be 13 billion light years from us. It exists in a much older part of the universe.

(credit: NASA, ESA, G.

Illingworth (University of
California, Santa Cruz),
R. Bouwens (University
of California, Santa Cruz
and Leiden University),
and the HUDF09 Team)

Consider the fact that the light we receive from these vast distances has been on its way to us for a long time. In fact, the time in years is the same as the distance in light years. For example, the Andromeda galaxy is 2 million light years away, so that the light now reaching us left it 2 million years ago. If we could be there now, Andromeda would be different. Similarly, light from the most distant galaxy left it 14 billion years ago. We have an incredible view of the past when looking great distances. We can try to see if the universe was different then—if distant galaxies are more tightly packed or have younger-looking stars, for example, than closer galaxies, in which case there has been an evolution in time. But the problem is that the uncertainties in our data are great. Cosmology is almost typified by these large uncertainties, so that we must be especially cautious in drawing conclusions. One consequence is that there are more questions than answers, and so there are many competing theories. Another consequence is that any hard data produce a major result. Discoveries of some importance are being made on a regular basis, the hallmark of a field in its golden age.

Perhaps the most important characteristic of the universe is that all galaxies except those in our local cluster seem to be moving away from us at speeds proportional to their distance from our galaxy. It looks as if a gigantic explosion, universally called the **Big Bang**, threw matter out some billions of years ago. This amazing conclusion is based on the pioneering work of Edwin Hubble (1889–1953), the American astronomer. In the 1920s, Hubble first demonstrated conclusively that other galaxies, many previously called nebulae or clouds of stars, were outside our own. He then found that all but the closest galaxies have a red shift in their hydrogen spectra that is proportional to their distance. The explanation is that there is a **cosmological red shift** due to the expansion of space itself. The photon

wavelength is stretched in transit from the source to the observer. Double the distance, and the red shift is doubled. While this cosmological red shift is often called a Doppler shift, it is not—space itself is expanding. There is no center of expansion in the universe. All observers see themselves as stationary; the other objects in space appear to be moving away from them. Hubble was directly responsible for discovering that the universe was much larger than had previously been imagined and that it had this amazing characteristic of rapid expansion.

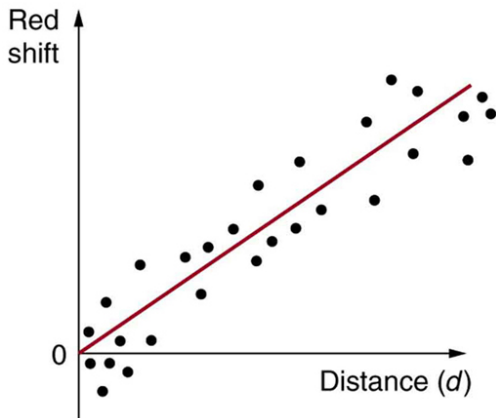
Universal expansion on the scale of galactic clusters (that is, galaxies at smaller distances are not uniformly receding from one another) is an integral part of modern cosmology. For galaxies farther away than about 50 Mly (50 million light years), the expansion is uniform with variations due to local motions of galaxies within clusters. A representative recession velocity v can be obtained from the simple formula

Equation:

$$v = H_0 d,$$

where d is the distance to the galaxy and H_0 is the **Hubble constant**. The Hubble constant is a central concept in cosmology. Its value is determined by taking the slope of a graph of velocity versus distance, obtained from red shift measurements, such as shown in [\[link\]](#). We shall use an approximate value of $H_0 = 20 \text{ km/s} \cdot \text{Mly}$. Thus, $v = H_0 d$ is an average behavior for all but the closest galaxies. For example, a galaxy 100 Mly away (as determined by its size and brightness) typically moves away from us at a speed of $v = (20 \text{ km/s} \cdot \text{Mly})(100 \text{ Mly}) = 2000 \text{ km/s}$. There can be variations in this speed due to so-called local motions or interactions with neighboring galaxies. Conversely, if a galaxy is found to be moving away from us at speed of 100,000 km/s based on its red shift, it is at a distance

$d = v/H_0 = (10,000 \text{ km/s})/(20 \text{ km/s} \cdot \text{Mly}) = 5000 \text{ Mly} = 5 \text{ Gly}$ or $5 \times 10^9 \text{ ly}$. This last calculation is approximate, because it assumes the expansion rate was the same 5 billion years ago as now. A similar calculation in Hubble's measurement changed the notion that the universe is in a steady state.

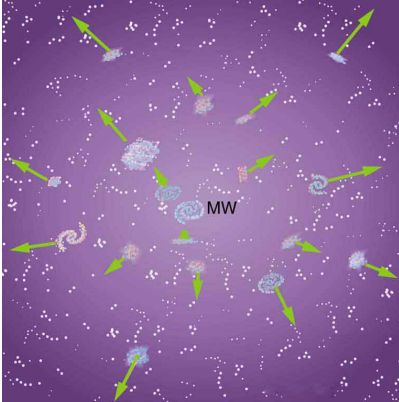


This graph of red shift versus distance for galaxies shows a linear relationship, with larger red shifts at greater distances, implying an expanding universe. The slope gives an approximate value for the expansion rate. (credit: John Cub).

One of the most intriguing developments recently has been the discovery that the expansion of the universe may be *faster now* than in the past, rather than slowing due to gravity as expected. Various groups have been looking, in particular, at supernovas in moderately distant galaxies (less than 1 Gly) to get improved distance measurements. Those distances are larger than expected for the observed galactic red shifts, implying the expansion was slower when that light was emitted. This has cosmological consequences that are discussed in [Dark Matter and Closure](#). The first results, published in 1999, are only the beginning of emerging data, with astronomy now entering a data-rich era.

[\[link\]](#) shows how the recession of galaxies looks like the remnants of a gigantic explosion, the famous Big Bang. Extrapolating backward in time,

the Big Bang would have occurred between 13 and 15 billion years ago when all matter would have been at a point. Questions instantly arise. What caused the explosion? What happened before the Big Bang? Was there a before, or did time start then? Will the universe expand forever, or will gravity reverse it into a Big Crunch? And is there other evidence of the Big Bang besides the well-documented red shifts?

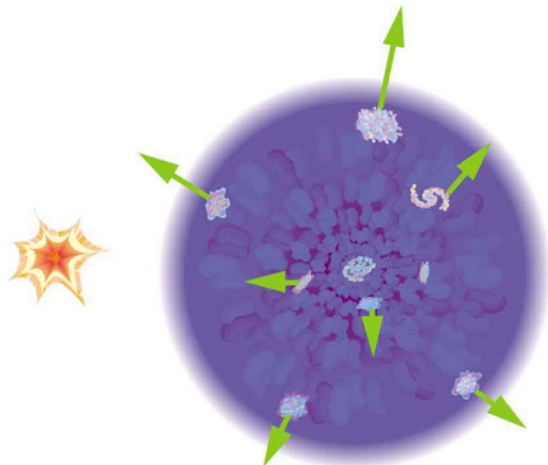


Galaxies are flying
apart from one
another, with the
more distant
moving faster as if
a primordial
explosion expelled
the matter from
which they formed.
The most distant
known galaxies
move nearly at the
speed of light
relative to us.

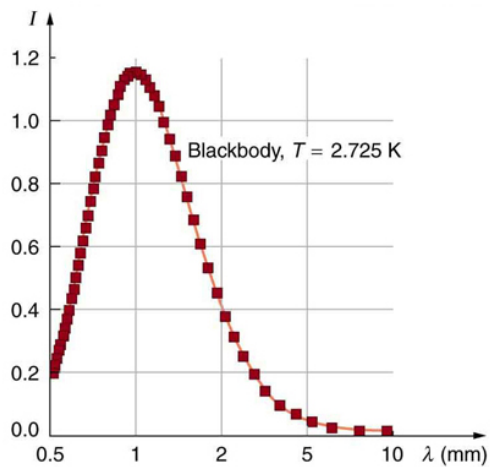
The Russian-born American physicist George Gamow (1904–1968) was among the first to note that, if there was a Big Bang, the remnants of the primordial fireball should still be evident and should be blackbody radiation. Since the radiation from this fireball has been traveling to us

since shortly after the Big Bang, its wavelengths should be greatly stretched. It will look as if the fireball has cooled in the billions of years since the Big Bang. Gamow and collaborators predicted in the late 1940s that there should be blackbody radiation from the explosion filling space with a characteristic temperature of about 7 K. Such blackbody radiation would have its peak intensity in the microwave part of the spectrum. (See [\[link\]](#).) In 1964, Arno Penzias and Robert Wilson, two American scientists working with Bell Telephone Laboratories on a low-noise radio antenna, detected the radiation and eventually recognized it for what it is.

[\[link\]](#)(b) shows the spectrum of this microwave radiation that permeates space and is of cosmic origin. It is the most perfect blackbody spectrum known, and the temperature of the fireball remnant is determined from it to be 2.725 ± 0.002 K. The detection of what is now called the **cosmic microwave background** (CMBR) was so important (generally considered as important as Hubble's detection that the galactic red shift is proportional to distance) that virtually every scientist has accepted the expansion of the universe as fact. Penzias and Wilson shared the 1978 Nobel Prize in Physics for their discovery.



(a)



(b)

(a) The Big Bang is used to explain the present observed expansion of the universe. It was an incredibly energetic explosion some 10 to 20 billion years ago. After expanding and cooling, galaxies form inside the now-cold remnants of the primordial fireball. (b) The spectrum of cosmic microwave radiation is the most perfect blackbody

spectrum ever detected. It is characteristic of a temperature of 2.725 K, the expansion-cooled temperature of the Big Bang's remnant. This radiation can be measured coming from any direction in space not obscured by some other source. It is compelling evidence of the creation of the universe in a gigantic explosion, already indicated by galactic red shifts.

Note:

Making Connections: Cosmology and Particle Physics

There are many connections of cosmology—by definition involving physics on the largest scale—with particle physics—by definition physics on the smallest scale. Among these are the dominance of matter over antimatter, the nearly perfect uniformity of the cosmic microwave background, and the mere existence of galaxies.

Matter versus antimatter

We know from direct observation that antimatter is rare. The Earth and the solar system are nearly pure matter. Space probes and cosmic rays give direct evidence—the landing of the Viking probes on Mars would have been spectacular explosions of mutual annihilation energy if Mars were antimatter. We also know that most of the universe is dominated by matter. This is proven by the lack of annihilation radiation coming to us from space, particularly the relative absence of 0.511-MeV γ rays created by the

mutual annihilation of electrons and positrons. It seemed possible that there could be entire solar systems or galaxies made of antimatter in perfect symmetry with our matter-dominated systems. But the interactions between stars and galaxies would sometimes bring matter and antimatter together in large amounts. The annihilation radiation they would produce is simply not observed. Antimatter in nature is created in particle collisions and in β^+ decays, but only in small amounts that quickly annihilate, leaving almost pure matter surviving.

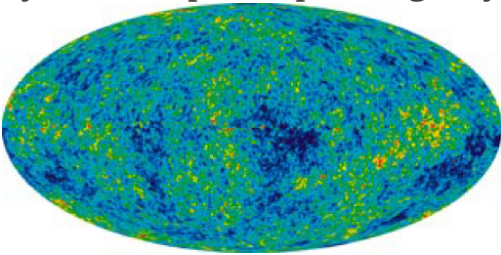
Particle physics seems symmetric in matter and antimatter. Why isn't the cosmos? The answer is that particle physics is not quite perfectly symmetric in this regard. The decay of one of the neutral K -mesons, for example, preferentially creates more matter than antimatter. This is caused by a fundamental small asymmetry in the basic forces. This small asymmetry produced slightly more matter than antimatter in the early universe. If there was only one part in 10^9 more matter (a small asymmetry), the rest would annihilate pair for pair, leaving nearly pure matter to form the stars and galaxies we see today. So the vast number of stars we observe may be only a tiny remnant of the original matter created in the Big Bang. Here at last we see a very real and important asymmetry in nature. Rather than be disturbed by an asymmetry, most physicists are impressed by how small it is. Furthermore, if the universe were completely symmetric, the mutual annihilation would be more complete, leaving far less matter to form us and the universe we know.

How can something so old have so few wrinkles?

A troubling aspect of cosmic microwave background radiation (CMBR) was soon recognized. True, the CMBR verified the Big Bang, had the correct temperature, and had a blackbody spectrum as expected. But the CMBR was *too* smooth—it looked identical in every direction. Galaxies and other similar entities could not be formed without the existence of fluctuations in the primordial stages of the universe and so there should be hot and cool spots in the CMBR, nicknamed wrinkles, corresponding to dense and sparse regions of gas caused by turbulence or early fluctuations. Over time, dense regions would contract under gravity and form stars and galaxies. Why aren't the fluctuations there? (This is a good example of an answer producing more questions.) Furthermore, galaxies are observed very

far from us, so that they formed very long ago. The problem was to explain how galaxies could form so early and so quickly after the Big Bang if its remnant fingerprint is perfectly smooth. The answer is that if you look very closely, the CMBR is not perfectly smooth, only extremely smooth.

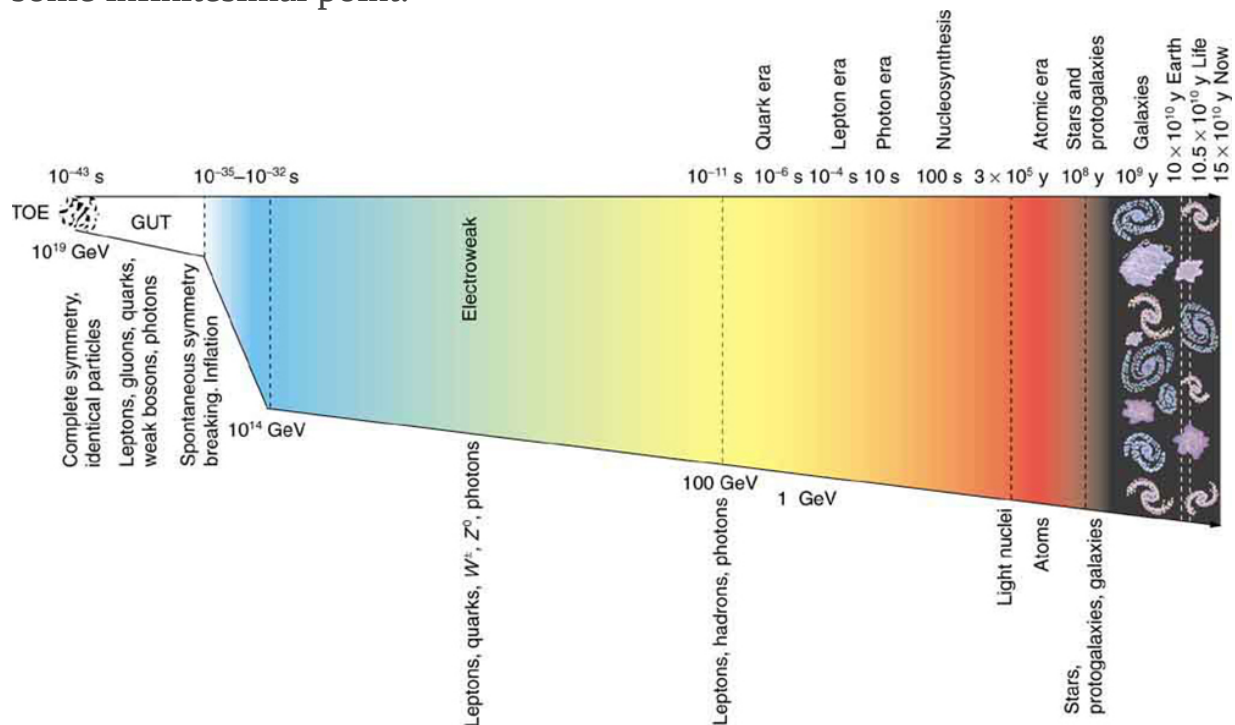
A satellite called the Cosmic Background Explorer (COBE) carried an instrument that made very sensitive and accurate measurements of the CMBR. In April of 1992, there was extraordinary publicity of COBE's first results—there were small fluctuations in the CMBR. Further measurements were carried out by experiments including NASA's Wilkinson Microwave Anisotropy Probe (WMAP), which launched in 2001. Data from WMAP provided a much more detailed picture of the CMBR fluctuations. (See [\[link\]](#).) These amount to temperature fluctuations of only $200\ \mu\text{K}$ out of $2.7\ \text{K}$, better than one part in 1000. The WMAP experiment will be followed up by the European Space Agency's Planck Surveyor, which launched in 2009.



This map of the sky uses color to show fluctuations, or wrinkles, in the cosmic microwave background observed with the WMAP spacecraft. The Milky Way has been removed for clarity. Red represents higher temperature and higher density, while blue is lower temperature and density. The fluctuations are small, less than one part in 1000, but these are still thought to be the

cause of the eventual
formation of galaxies.
(credit: NASA/WMAP
Science Team)

Let us now examine the various stages of the overall evolution of the universe from the Big Bang to the present, illustrated in [\[link\]](#). Note that scientific notation is used to encompass the many orders of magnitude in time, energy, temperature, and size of the universe. Going back in time, the two lines approach but do not cross (there is no zero on an exponential scale). Rather, they extend indefinitely in ever-smaller time intervals to some infinitesimal point.



The evolution of the universe from the Big Bang onward is intimately tied to the laws of physics, especially those of particle physics at the earliest stages. The universe is relativistic throughout its history. Theories of the unification of forces at high energies may be verified by their shaping of the universe and its evolution.

Going back in time is equivalent to what would happen if expansion stopped and gravity pulled all the galaxies together, compressing and heating all matter. At a time long ago, the temperature and density were too high for stars and galaxies to exist. Before then, there was a time when the temperature was too great for atoms to exist. And farther back yet, there was a time when the temperature and density were so great that nuclei could not exist. Even farther back in time, the temperature was so high that average kinetic energy was great enough to create short-lived particles, and the density was high enough to make this likely. When we extrapolate back to the point of W^\pm and Z^0 production (thermal energies reaching 1 TeV, or a temperature of about 10^{15} K), we reach the limits of what we know directly about particle physics. This is at a time about 10^{-12} s after the Big Bang. While 10^{-12} s may seem to be negligibly close to the instant of creation, it is not. There are important stages before this time that are tied to the unification of forces. At those stages, the universe was at extremely high energies and average particle separations were smaller than we can achieve with accelerators. What happened in the early stages before 10^{-12} s is crucial to all later stages and is possibly discerned by observing present conditions in the universe. One of these is the smoothness of the CMBR.

Names are given to early stages representing key conditions. The stage before 10^{-11} s back to 10^{-34} s is called the **electroweak epoch**, because the electromagnetic and weak forces become identical for energies above about 100 GeV. As discussed earlier, theorists expect that the strong force becomes identical to and thus unified with the electroweak force at energies of about 10^{14} GeV. The average particle energy would be this great at 10^{-34} s after the Big Bang, if there are no surprises in the unknown physics at energies above about 1 TeV. At the immense energy of 10^{14} GeV (corresponding to a temperature of about 10^{26} K), the W^\pm and Z^0 carrier particles would be transformed into massless gauge bosons to accomplish the unification. Before 10^{-34} s back to about 10^{-43} s, we have Grand Unification in the **GUT epoch**, in which all forces except gravity are identical. At 10^{-43} s, the average energy reaches the immense 10^{19} GeV needed to unify gravity with the other forces in TOE, the Theory of Everything. Before that time is the **TOE epoch**, but we have almost no idea

as to the nature of the universe then, since we have no workable theory of quantum gravity. We call the hypothetical unified force **superforce**.

Now let us imagine starting at TOE and moving forward in time to see what type of universe is created from various events along the way. As temperatures and average energies decrease with expansion, the universe reaches the stage where average particle separations are large enough to see differences between the strong and electroweak forces (at about 10^{-35} s). After this time, the forces become distinct in almost all interactions—they are no longer unified or symmetric. This transition from GUT to electroweak is an example of **spontaneous symmetry breaking**, in which conditions spontaneously evolved to a point where the forces were no longer unified, breaking that symmetry. This is analogous to a phase transition in the universe, and a clever proposal by American physicist Alan Guth in the early 1980s ties it to the smoothness of the CMBR. Guth proposed that spontaneous symmetry breaking (like a phase transition during cooling of normal matter) released an immense amount of energy that caused the universe to expand extremely rapidly for the brief time from 10^{-35} s to about 10^{-32} s. This expansion may have been by an incredible factor of 10^{50} or more in the size of the universe and is thus called the **inflationary scenario**. One result of this inflation is that it would stretch the wrinkles in the universe nearly flat, leaving an extremely smooth CMBR. While speculative, there is as yet no other plausible explanation for the smoothness of the CMBR. Unless the CMBR is not really cosmic but local in origin, the distances between regions of similar temperatures are too great for any coordination to have caused them, since any coordination mechanism must travel at the speed of light. Again, particle physics and cosmology are intimately entwined. There is little hope that we may be able to test the inflationary scenario directly, since it occurs at energies near 10^{14} GeV, vastly greater than the limits of modern accelerators. But the idea is so attractive that it is incorporated into most cosmological theories.

Characteristics of the present universe may help us determine the validity of this intriguing idea. Additionally, the recent indications that the universe's expansion rate may be *increasing* (see [Dark Matter and Closure](#)) could even imply that we are *in* another inflationary epoch.

It is important to note that, if conditions such as those found in the early universe could be created in the laboratory, we would see the unification of forces directly today. The forces have not changed in time, but the average energy and separation of particles in the universe have. As discussed in [The Four Basic Forces](#), the four basic forces in nature are distinct under most circumstances found today. The early universe and its remnants provide evidence from times when they were unified under most circumstances.

Section Summary

- Cosmology is the study of the character and evolution of the universe.
- The two most important features of the universe are the cosmological red shifts of its galaxies being proportional to distance and its cosmic microwave background (CMBR). Both support the notion that there was a gigantic explosion, known as the Big Bang that created the universe.
- Galaxies farther away than our local group have, on an average, a recessional velocity given by

Equation:

$$v = H_0 d,$$

where d is the distance to the galaxy and H_0 is the Hubble constant, taken to have the average value $H_0 = 20 \text{ km/s} \cdot \text{Mly}$.

- Explanations of the large-scale characteristics of the universe are intimately tied to particle physics.
- The dominance of matter over antimatter and the smoothness of the CMBR are two characteristics that are tied to particle physics.
- The epochs of the universe are known back to very shortly after the Big Bang, based on known laws of physics.
- The earliest epochs are tied to the unification of forces, with the electroweak epoch being partially understood, the GUT epoch being speculative, and the TOE epoch being highly speculative since it involves an unknown single superforce.
- The transition from GUT to electroweak is called spontaneous symmetry breaking. It released energy that caused the inflationary

scenario, which in turn explains the smoothness of the CMBR.

Conceptual Questions

Exercise:

Problem:

Explain why it only *appears* that we are at the center of expansion of the universe and why an observer in another galaxy would see the same relative motion of all but the closest galaxies away from her.

Exercise:

Problem:

If there is no observable edge to the universe, can we determine where its center of expansion is? Explain.

Exercise:

Problem: If the universe is infinite, does it have a center? Discuss.

Exercise:

Problem:

Another known cause of red shift in light is the source being in a high gravitational field. Discuss how this can be eliminated as the source of galactic red shifts, given that the shifts are proportional to distance and not to the size of the galaxy.

Exercise:

Problem:

If some unknown cause of red shift—such as light becoming “tired” from traveling long distances through empty space—is discovered, what effect would there be on cosmology?

Exercise:

Problem:

Olbers's paradox poses an interesting question: If the universe is infinite, then any line of sight should eventually fall on a star's surface. Why then is the sky dark at night? Discuss the commonly accepted evolution of the universe as a solution to this paradox.

Exercise:**Problem:**

If the cosmic microwave background radiation (CMBR) is the remnant of the Big Bang's fireball, we expect to see hot and cold regions in it. What are two causes of these wrinkles in the CMBR? Are the observed temperature variations greater or less than originally expected?

Exercise:**Problem:**

The decay of one type of K -meson is cited as evidence that nature favors matter over antimatter. Since mesons are composed of a quark and an antiquark, is it surprising that they would preferentially decay to one type over another? Is this an asymmetry in nature? Is the predominance of matter over antimatter an asymmetry?

Exercise:**Problem:**

Distances to local galaxies are determined by measuring the brightness of stars, called Cepheid variables, that can be observed individually and that have absolute brightnesses at a standard distance that are well known. Explain how the measured brightness would vary with distance as compared with the absolute brightness.

Exercise:

Problem:

Distances to very remote galaxies are estimated based on their apparent type, which indicate the number of stars in the galaxy, and their measured brightness. Explain how the measured brightness would vary with distance. Would there be any correction necessary to compensate for the red shift of the galaxy (all distant galaxies have significant red shifts)? Discuss possible causes of uncertainties in these measurements.

Exercise:**Problem:**

If the smallest meaningful time interval is greater than zero, will the lines in [\[link\]](#) ever meet?

Problems & Exercises**Exercise:****Problem:**

Find the approximate mass of the luminous matter in the Milky Way galaxy, given it has approximately 10^{11} stars of average mass 1.5 times that of our Sun.

Solution:

$$3 \times 10^{41} \text{ kg}$$

Exercise:

Problem:

Find the approximate mass of the dark and luminous matter in the Milky Way galaxy. Assume the luminous matter is due to approximately 10^{11} stars of average mass 1.5 times that of our Sun, and take the dark matter to be 10 times as massive as the luminous matter.

Exercise:**Problem:**

(a) Estimate the mass of the luminous matter in the known universe, given there are 10^{11} galaxies, each containing 10^{11} stars of average mass 1.5 times that of our Sun. (b) How many protons (the most abundant nuclide) are there in this mass? (c) Estimate the total number of particles in the observable universe by multiplying the answer to (b) by two, since there is an electron for each proton, and then by 10^9 , since there are far more particles (such as photons and neutrinos) in space than in luminous matter.

Solution:

(a) $3 \times 10^{52} \text{ kg}$

(b) 2×10^{79}

(c) 4×10^{88}

Exercise:**Problem:**

If a galaxy is 500 Mly away from us, how fast do we expect it to be moving and in what direction?

Exercise:

Problem:

On average, how far away are galaxies that are moving away from us at 2.0% of the speed of light?

Solution:

0.30 Gly

Exercise:**Problem:**

Our solar system orbits the center of the Milky Way galaxy. Assuming a circular orbit 30,000 ly in radius and an orbital speed of 250 km/s, how many years does it take for one revolution? Note that this is approximate, assuming constant speed and circular orbit, but it is representative of the time for our system and local stars to make one revolution around the galaxy.

Exercise:**Problem:**

(a) What is the approximate speed relative to us of a galaxy near the edge of the known universe, some 10 Gly away? (b) What fraction of the speed of light is this? Note that we have observed galaxies moving away from us at greater than $0.9c$.

Solution:

(a) 2.0×10^5 km/s

(b) $0.67c$

Exercise:

Problem:

(a) Calculate the approximate age of the universe from the average value of the Hubble constant, $H_0 = 20 \text{ km/s} \cdot \text{Mly}$. To do this, calculate the time it would take to travel 1 Mly at a constant expansion rate of 20 km/s. (b) If deceleration is taken into account, would the actual age of the universe be greater or less than that found here? Explain.

Exercise:**Problem:**

Assuming a circular orbit for the Sun about the center of the Milky Way galaxy, calculate its orbital speed using the following information: The mass of the galaxy is equivalent to a single mass 1.5×10^{11} times that of the Sun (or $3 \times 10^{41} \text{ kg}$), located 30,000 ly away.

Solution:

$$2.7 \times 10^5 \text{ m/s}$$

Exercise:**Problem:**

(a) What is the approximate force of gravity on a 70-kg person due to the Andromeda galaxy, assuming its total mass is 10^{13} that of our Sun and acts like a single mass 2 Mly away? (b) What is the ratio of this force to the person's weight? Note that Andromeda is the closest large galaxy.

Exercise:**Problem:**

Andromeda galaxy is the closest large galaxy and is visible to the naked eye. Estimate its brightness relative to the Sun, assuming it has luminosity 10^{12} times that of the Sun and lies 2 Mly away.

Solution:

6×10^{-11} (an overestimate, since some of the light from Andromeda is blocked by gas and dust within that galaxy)

Exercise:**Problem:**

(a) A particle and its antiparticle are at rest relative to an observer and annihilate (completely destroying both masses), creating two γ rays of equal energy. What is the characteristic γ -ray energy you would look for if searching for evidence of proton-antiproton annihilation? (The fact that such radiation is rarely observed is evidence that there is very little antimatter in the universe.) (b) How does this compare with the 0.511-MeV energy associated with electron-positron annihilation?

Exercise:**Problem:**

The average particle energy needed to observe unification of forces is estimated to be 10^{19} GeV. (a) What is the rest mass in kilograms of a particle that has a rest mass of 10^{19} GeV/ c^2 ? (b) How many times the mass of a hydrogen atom is this?

Solution:

(a) 2×10^{-8} kg

(b) 1×10^{19}

Exercise:

Problem:

The peak intensity of the CMBR occurs at a wavelength of 1.1 mm. (a) What is the energy in eV of a 1.1-mm photon? (b) There are approximately 10^9 photons for each massive particle in deep space. Calculate the energy of 10^9 such photons. (c) If the average massive particle in space has a mass half that of a proton, what energy would be created by converting its mass to energy? (d) Does this imply that space is “matter dominated”? Explain briefly.

Exercise:**Problem:**

(a) What Hubble constant corresponds to an approximate age of the universe of 10^{10} y? To get an approximate value, assume the expansion rate is constant and calculate the speed at which two galaxies must move apart to be separated by 1 Mly (present average galactic separation) in a time of 10^{10} y. (b) Similarly, what Hubble constant corresponds to a universe approximately 2×10^{10} -y old?

Solution:

(a) $30 \text{ km/s} \cdot \text{Mly}$

(b) $15 \text{ km/s} \cdot \text{Mly}$

Exercise:**Problem:**

Show that the velocity of a star orbiting its galaxy in a circular orbit is inversely proportional to the square root of its orbital radius, assuming the mass of the stars inside its orbit acts like a single mass at the center of the galaxy. You may use an equation from a previous chapter to support your conclusion, but you must justify its use and define all terms used.

Exercise:

Problem:

The core of a star collapses during a supernova, forming a neutron star. Angular momentum of the core is conserved, and so the neutron star spins rapidly. If the initial core radius is 5.0×10^5 km and it collapses to 10.0 km, find the neutron star's angular velocity in revolutions per second, given the core's angular velocity was originally 1 revolution per 30.0 days.

Solution:

960 rev/s

Exercise:**Problem:**

Using data from the previous problem, find the increase in rotational kinetic energy, given the core's mass is 1.3 times that of our Sun. Where does this increase in kinetic energy come from?

Exercise:**Problem:**

Distances to the nearest stars (up to 500 ly away) can be measured by a technique called parallax, as shown in [\[link\]](#). What are the angles θ_1 and θ_2 relative to the plane of the Earth's orbit for a star 4.0 ly directly above the Sun?

Solution:

89.999773° (many digits are used to show the difference between 90°)

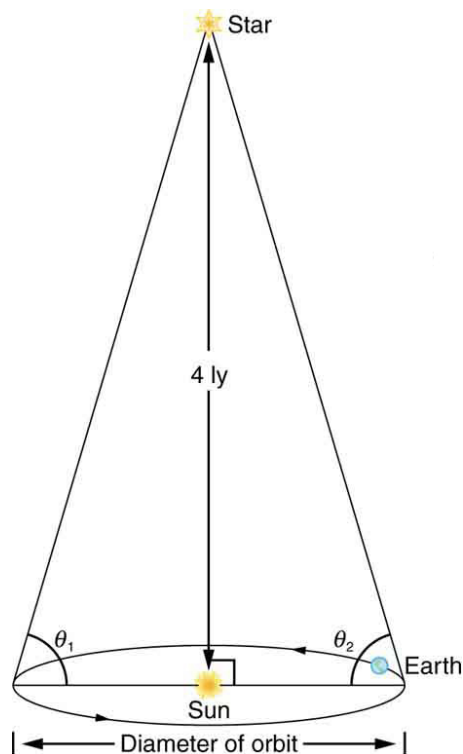
Exercise:

Problem:

(a) Use the Heisenberg uncertainty principle to calculate the uncertainty in energy for a corresponding time interval of 10^{-43} s. (b) Compare this energy with the 10^{19} GeV unification-of-forces energy and discuss why they are similar.

Exercise:**Problem: Construct Your Own Problem**

Consider a star moving in a circular orbit at the edge of a galaxy. Construct a problem in which you calculate the mass of that galaxy in kg and in multiples of the solar mass based on the velocity of the star and its distance from the center of the galaxy.



Distances to nearby stars are measured using triangulation, also called the parallax method. The angle of

line of sight to the star
is measured at
intervals six months
apart, and the distance
is calculated by using
the known diameter of
the Earth's orbit. This
can be done for stars
up to about 500 ly
away.

Glossary

Big Bang

a gigantic explosion that threw out matter a few billion years ago

cosmic microwave background

the spectrum of microwave radiation of cosmic origin

cosmological red shift

the photon wavelength is stretched in transit from the source to the observer because of the expansion of space itself

cosmology

the study of the character and evolution of the universe

electroweak epoch

the stage before 10^{-11} back to 10^{-34} after the Big Bang

GUT epoch

the time period from 10^{-43} to 10^{-34} after the Big Bang, when Grand Unification Theory, in which all forces except gravity are identical, governed the universe

Hubble constant

a central concept in cosmology whose value is determined by taking the slope of a graph of velocity versus distance, obtained from red shift measurements

inflationary scenario

the rapid expansion of the universe by an incredible factor of 10^{-50} for the brief time from 10^{-35} to about 10^{-32} s

spontaneous symmetry breaking

the transition from GUT to electroweak where the forces were no longer unified

superforce

hypothetical unified force in TOE epoch

TOE epoch

before 10^{-43} after the Big Bang

General Relativity and Quantum Gravity

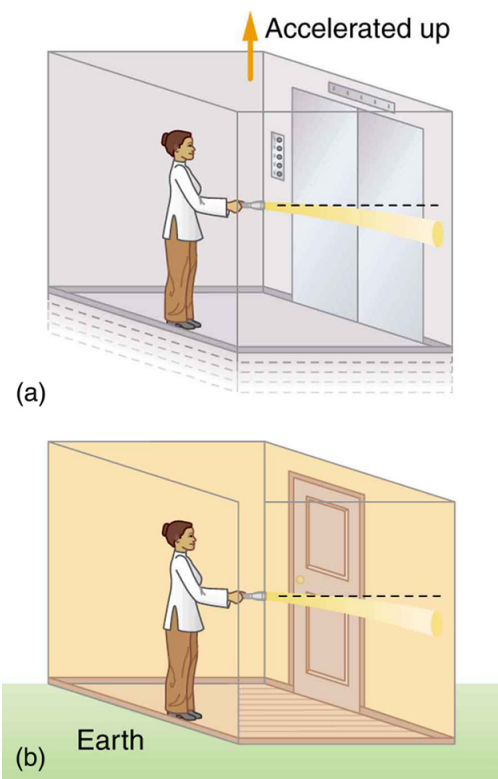
- Explain the effect of gravity on light.
- Discuss black hole.
- Explain quantum gravity.

When we talk of black holes or the unification of forces, we are actually discussing aspects of general relativity and quantum gravity. We know from [Special Relativity](#) that relativity is the study of how different observers measure the same event, particularly if they move relative to one another. Einstein's theory of **general relativity** describes all types of relative motion including accelerated motion and the effects of gravity. General relativity encompasses special relativity and classical relativity in situations where acceleration is zero and relative velocity is small compared with the speed of light. Many aspects of general relativity have been verified experimentally, some of which are better than science fiction in that they are bizarre but true. **Quantum gravity** is the theory that deals with particle exchange of gravitons as the mechanism for the force, and with extreme conditions where quantum mechanics and general relativity must both be used. A good theory of quantum gravity does not yet exist, but one will be needed to understand how all four forces may be unified. If we are successful, the theory of quantum gravity will encompass all others, from classical physics to relativity to quantum mechanics—truly a Theory of Everything (TOE).

General Relativity

Einstein first considered the case of no observer acceleration when he developed the revolutionary special theory of relativity, publishing his first work on it in 1905. By 1916, he had laid the foundation of general relativity, again almost on his own. Much of what Einstein did to develop his ideas was to mentally analyze certain carefully and clearly defined situations—doing this is to perform a **thought experiment**. [\[link\]](#) illustrates a thought experiment like the ones that convinced Einstein that light must fall in a gravitational field. Think about what a person feels in an elevator that is accelerated upward. It is identical to being in a stationary elevator in a gravitational field. The feet of a person are pressed against the floor, and

objects released from hand fall with identical accelerations. In fact, it is not possible, without looking outside, to know what is happening—acceleration upward or gravity. This led Einstein to correctly postulate that acceleration and gravity will produce identical effects in all situations. So, if acceleration affects light, then gravity will, too. [\[link\]](#) shows the effect of acceleration on a beam of light shone horizontally at one wall. Since the accelerated elevator moves up during the time light travels across the elevator, the beam of light strikes low, seeming to the person to bend down. (Normally a tiny effect, since the speed of light is so great.) The same effect must occur due to gravity, Einstein reasoned, since there is no way to tell the effects of gravity acting downward from acceleration of the elevator upward. Thus gravity affects the path of light, even though we think of gravity as acting between masses and photons are massless.



(a) A beam of light emerges from a flashlight in an upward-accelerating

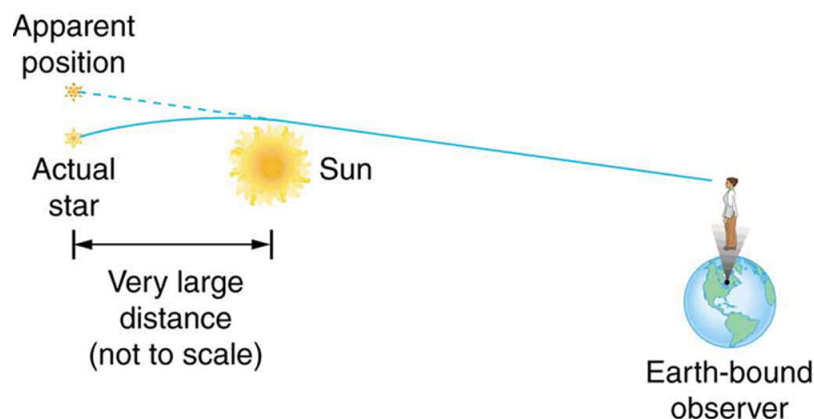
elevator. Since the elevator moves up during the time the light takes to reach the wall, the beam strikes lower than it would if the elevator were not accelerated. (b) Gravity has the same effect on light, since it is not possible to tell whether the elevator is accelerating upward or acted upon by gravity.

Einstein's theory of general relativity got its first verification in 1919 when starlight passing near the Sun was observed during a solar eclipse. (See [\[link\]](#).) During an eclipse, the sky is darkened and we can briefly see stars. Those in a line of sight nearest the Sun should have a shift in their apparent positions. Not only was this shift observed, but it agreed with Einstein's predictions well within experimental uncertainties. This discovery created a scientific and public sensation. Einstein was now a folk hero as well as a very great scientist. The bending of light by matter is equivalent to a bending of space itself, with light following the curve. This is another radical change in our concept of space and time. It is also another connection that any particle with mass or energy (massless photons) is affected by gravity.

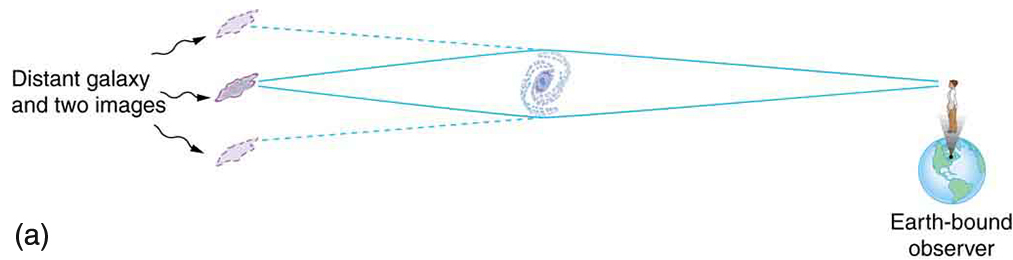
There are several current forefront efforts related to general relativity. One is the observation and analysis of gravitational lensing of light. Another is analysis of the definitive proof of the existence of black holes. Direct observation of gravitational waves or moving wrinkles in space is being searched for. Theoretical efforts are also being aimed at the possibility of time travel and wormholes into other parts of space due to black holes.

Gravitational lensing

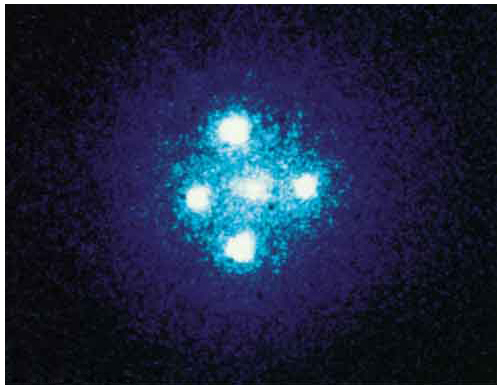
As you can see in [\[link\]](#), light is bent toward a mass, producing an effect much like a converging lens (large masses are needed to produce observable effects). On a galactic scale, the light from a distant galaxy could be “lensed” into several images when passing close by another galaxy on its way to Earth. Einstein predicted this effect, but he considered it unlikely that we would ever observe it. A number of cases of this effect have now been observed; one is shown in [\[link\]](#). This effect is a much larger scale verification of general relativity. But such gravitational lensing is also useful in verifying that the red shift is proportional to distance. The red shift of the intervening galaxy is always less than that of the one being lensed, and each image of the lensed galaxy has the same red shift. This verification supplies more evidence that red shift is proportional to distance. Confidence that the multiple images are not different objects is bolstered by the observations that if one image varies in brightness over time, the others also vary in the same manner.



This schematic shows how light passing near a massive body like the Sun is curved toward it. The light that reaches the Earth then seems to be coming from different locations than the known positions of the originating stars. Not only was this effect observed, the amount of bending was precisely what Einstein predicted in his general theory of relativity.



(a)



(b)

(a) Light from a distant galaxy can travel different paths to the Earth because it is bent around an intermediary galaxy by gravity. This produces several images of the more distant galaxy. (b) The images around the central galaxy are produced by gravitational lensing. Each image has the same spectrum and a larger red shift than the intermediary. (credit: NASA, ESA, and STScI)

Black holes

Black holes are objects having such large gravitational fields that things can fall in, but nothing, not even light, can escape. Bodies, like the Earth or the Sun, have what is called an **escape velocity**. If an object moves straight up from the body, starting at the escape velocity, it will just be able to escape the gravity of the body. The greater the acceleration of gravity on the body, the greater is the escape velocity. As long ago as the late 1700s, it was proposed that if the escape velocity is greater than the speed of light, then

light cannot escape. Simon Laplace (1749–1827), the French astronomer and mathematician, even incorporated this idea of a dark star into his writings. But the idea was dropped after Young’s double slit experiment showed light to be a wave. For some time, light was thought not to have particle characteristics and, thus, could not be acted upon by gravity. The idea of a black hole was very quickly reincarnated in 1916 after Einstein’s theory of general relativity was published. It is now thought that black holes can form in the supernova collapse of a massive star, forming an object perhaps 10 km across and having a mass greater than that of our Sun. It is interesting that several prominent physicists who worked on the concept, including Einstein, firmly believed that nature would find a way to prohibit such objects.

Black holes are difficult to observe directly, because they are small and no light comes directly from them. In fact, no light comes from inside the **event horizon**, which is defined to be at a distance from the object at which the escape velocity is exactly the speed of light. The radius of the event horizon is known as the **Schwarzschild radius** R_S and is given by

Equation:

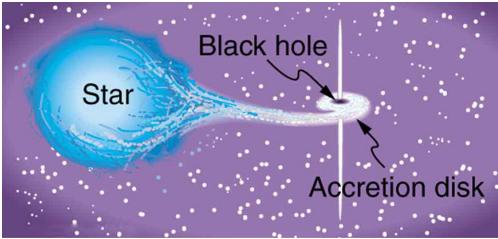
$$R_S = \frac{2GM}{c^2},$$

where G is the universal gravitational constant, M is the mass of the body, and c is the speed of light. The event horizon is the edge of the black hole and R_S is its radius (that is, the size of a black hole is twice R_S). Since G is small and c^2 is large, you can see that black holes are extremely small, only a few kilometers for masses a little greater than the Sun’s. The object itself is inside the event horizon.

Physics near a black hole is fascinating. Gravity increases so rapidly that, as you approach a black hole, the tidal effects tear matter apart, with matter closer to the hole being pulled in with much more force than that only slightly farther away. This can pull a companion star apart and heat inflowing gases to the point of producing X rays. (See [\[link\]](#).) We have observed X rays from certain binary star systems that are consistent with such a picture. This is not quite proof of black holes, because the X rays

could also be caused by matter falling onto a neutron star. These objects were first discovered in 1967 by the British astrophysicists, Jocelyn Bell and Anthony Hewish. **Neutron stars** are literally a star composed of neutrons. They are formed by the collapse of a star's core in a supernova, during which electrons and protons are forced together to form neutrons (the reverse of neutron β decay). Neutron stars are slightly larger than a black hole of the same mass and will not collapse further because of resistance by the strong force. However, neutron stars cannot have a mass greater than about eight solar masses or they must collapse to a black hole. With recent improvements in our ability to resolve small details, such as with the orbiting Chandra X-ray Observatory, it has become possible to measure the masses of X-ray-emitting objects by observing the motion of companion stars and other matter in their vicinity. What has emerged is a plethora of X-ray-emitting objects too massive to be neutron stars. This evidence is considered conclusive and the existence of black holes is widely accepted. These black holes are concentrated near galactic centers.

We also have evidence that supermassive black holes may exist at the cores of many galaxies, including the Milky Way. Such a black hole might have a mass millions or even billions of times that of the Sun, and it would probably have formed when matter first coalesced into a galaxy billions of years ago. Supporting this is the fact that very distant galaxies are more likely to have abnormally energetic cores. Some of the moderately distant galaxies, and hence among the younger, are known as **quasars** and emit as much or more energy than a normal galaxy but from a region less than a light year across. Quasar energy outputs may vary in times less than a year, so that the energy-emitting region must be less than a light year across. The best explanation of quasars is that they are young galaxies with a supermassive black hole forming at their core, and that they become less energetic over billions of years. In closer superactive galaxies, we observe tremendous amounts of energy being emitted from very small regions of space, consistent with stars falling into a black hole at the rate of one or more a month. The Hubble Space Telescope (1994) observed an accretion disk in the galaxy M87 rotating rapidly around a region of extreme energy emission. (See [\[link\]](#).) A jet of material being ejected perpendicular to the plane of rotation gives further evidence of a supermassive black hole as the engine.



A black hole is shown pulling matter away from a companion star, forming a superheated accretion disk where X rays are emitted before the matter disappears forever into the hole. The in-fall energy also ejects some material, forming the two vertical spikes. (See also the photograph in [Introduction to Frontiers of Physics](#).) There are several X-ray-emitting objects in space that are consistent with this picture and are likely to be black holes.

Gravitational waves

If a massive object distorts the space around it, like the foot of a water bug on the surface of a pond, then movement of the massive object should create waves in space like those on a pond. **Gravitational waves** are mass-created distortions in space that propagate at the speed of light and are predicted by general relativity. Since gravity is by far the weakest force, extreme conditions are needed to generate significant gravitational waves. Gravity near binary neutron star systems is so great that significant gravitational wave energy is radiated as the two neutron stars orbit one

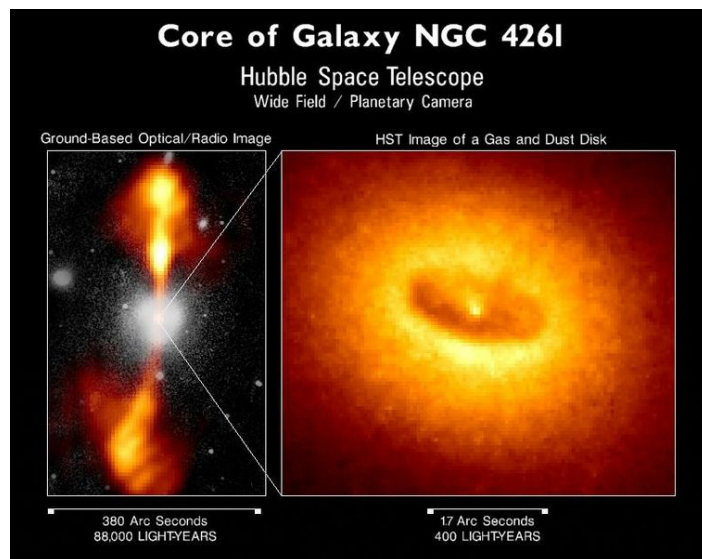
another. American astronomers, Joseph Taylor and Russell Hulse, measured changes in the orbit of such a binary neutron star system. They found its orbit to change precisely as predicted by general relativity, a strong indication of gravitational waves, and were awarded the 1993 Nobel Prize. But direct detection of gravitational waves on Earth would be conclusive. For many years, various attempts have been made to detect gravitational waves by observing vibrations induced in matter distorted by these waves. American physicist Joseph Weber pioneered this field in the 1960s, but no conclusive events have been observed. (No gravity wave detectors were in operation at the time of the 1987A supernova, unfortunately.) There are now several ambitious systems of gravitational wave detectors in use around the world. These include the LIGO (Laser Interferometer Gravitational Wave Observatory) system with two laser interferometer detectors, one in the state of Washington and another in Louisiana (See [\[link\]](#)) and the VIRGO (Variability of Irradiance and Gravitational Oscillations) facility in Italy with a single detector.

Quantum Gravity

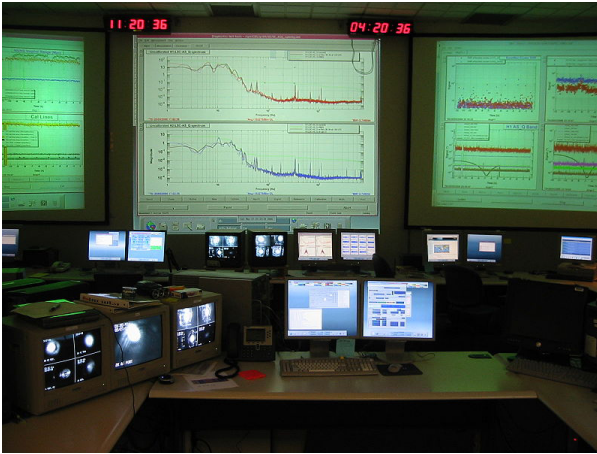
Black holes radiate

Quantum gravity is important in those situations where gravity is so extremely strong that it has effects on the quantum scale, where the other forces are ordinarily much stronger. The early universe was such a place, but black holes are another. The first significant connection between gravity and quantum effects was made by the Russian physicist Yakov Zel'dovich in 1971, and other significant advances followed from the British physicist Stephen Hawking. (See [\[link\]](#).) These two showed that black holes could radiate away energy by quantum effects just outside the event horizon (nothing can escape from inside the event horizon). Black holes are, thus, expected to radiate energy and shrink to nothing, although extremely slowly for most black holes. The mechanism is the creation of a particle-antiparticle pair from energy in the extremely strong gravitational field near the event horizon. One member of the pair falls into the hole and the other escapes, conserving momentum. (See [\[link\]](#).) When a black hole loses energy and, hence, rest mass, its event horizon shrinks, creating an even greater gravitational field. This increases the rate of pair production so that the process grows exponentially until the black hole is nuclear in size. A

final burst of particles and γ rays ensues. This is an extremely slow process for black holes about the mass of the Sun (produced by supernovas) or larger ones (like those thought to be at galactic centers), taking on the order of 10^{67} years or longer! Smaller black holes would evaporate faster, but they are only speculated to exist as remnants of the Big Bang. Searches for characteristic γ -ray bursts have produced events attributable to more mundane objects like neutron stars accreting matter.



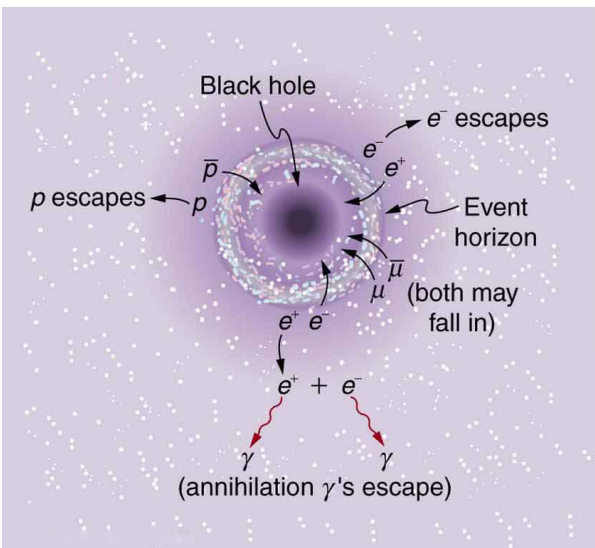
This Hubble Space Telescope photograph shows the extremely energetic core of the NGC 4261 galaxy. With the superior resolution of the orbiting telescope, it has been possible to observe the rotation of an accretion disk around the energy-producing object as well as to map jets of material being ejected from the object. A supermassive black hole is consistent with these observations, but other possibilities are not quite eliminated. (credit: NASA and ESA)



The control room of the LIGO gravitational wave detector. Gravitational waves will cause extremely small vibrations in a mass in this detector, which will be detected by laser interferometer techniques. Such detection in coincidence with other detectors and with astronomical events, such as supernovas, would provide direct evidence of gravitational waves. (credit: Tobin Fricke)



Stephen Hawking (b. 1942) has made many contributions to the theory of quantum gravity. Hawking is a long-time survivor of ALS and has produced popular books on general relativity, cosmology, and quantum gravity. (credit: Lwp Kommunikáció)



Gravity and quantum mechanics

come into play when a black hole creates a particle-antiparticle pair from the energy in its gravitational field. One member of the pair falls into the hole while the other escapes, removing energy and shrinking the black hole. The search is on for the characteristic energy.

Wormholes and time travel

The subject of time travel captures the imagination. Theoretical physicists, such as the American Kip Thorne, have treated the subject seriously, looking into the possibility that falling into a black hole could result in popping up in another time and place—a trip through a so-called wormhole. Time travel and wormholes appear in innumerable science fiction dramatizations, but the consensus is that time travel is not possible in theory. While still debated, it appears that quantum gravity effects inside a black hole prevent time travel due to the creation of particle pairs. Direct evidence is elusive.

The shortest time

Theoretical studies indicate that, at extremely high energies and correspondingly early in the universe, quantum fluctuations may make time intervals meaningful only down to some finite time limit. Early work indicated that this might be the case for times as long as 10^{-43} s, the time at which all forces were unified. If so, then it would be meaningless to consider the universe at times earlier than this. Subsequent studies indicate that the crucial time may be as short as 10^{-95} s. But the point remains—quantum gravity seems to imply that there is no such thing as a vanishingly short time. Time may, in fact, be grainy with no meaning to time intervals shorter than some tiny but finite size.

The future of quantum gravity

Not only is quantum gravity in its infancy, no one knows how to get started on a theory of gravitons and unification of forces. The energies at which TOE should be valid may be so high (at least 10^{19} GeV) and the necessary particle separation so small (less than 10^{-35} m) that only indirect evidence can provide clues. For some time, the common lament of theoretical physicists was one so familiar to struggling students—how do you even get started? But Hawking and others have made a start, and the approach many theorists have taken is called Superstring theory, the topic of the [Superstrings](#).

Section Summary

- Einstein's theory of general relativity includes accelerated frames and, thus, encompasses special relativity and gravity. Created by use of careful thought experiments, it has been repeatedly verified by real experiments.
- One direct result of this behavior of nature is the gravitational lensing of light by massive objects, such as galaxies, also seen in the microlensing of light by smaller bodies in our galaxy.
- Another prediction is the existence of black holes, objects for which the escape velocity is greater than the speed of light and from which nothing can escape.
- The event horizon is the distance from the object at which the escape velocity equals the speed of light c . It is called the Schwarzschild radius R_S and is given by

Equation:

$$R_S = \frac{2GM}{c^2},$$

where G is the universal gravitational constant, and M is the mass of the body.

- Physics is unknown inside the event horizon, and the possibility of wormholes and time travel are being studied.
- Candidates for black holes may power the extremely energetic emissions of quasars, distant objects that seem to be early stages of

galactic evolution.

- Neutron stars are stellar remnants, having the density of a nucleus, that hint that black holes could form from supernovas, too.
- Gravitational waves are wrinkles in space, predicted by general relativity but not yet observed, caused by changes in very massive objects.
- Quantum gravity is an incompletely developed theory that strives to include general relativity, quantum mechanics, and unification of forces (thus, a TOE).
- One unconfirmed connection between general relativity and quantum mechanics is the prediction of characteristic radiation from just outside black holes.

Conceptual Questions

Exercise:

Problem:

Quantum gravity, if developed, would be an improvement on both general relativity and quantum mechanics, but more mathematically difficult. Under what circumstances would it be necessary to use quantum gravity? Similarly, under what circumstances could general relativity be used? When could special relativity, quantum mechanics, or classical physics be used?

Exercise:

Problem:

Does observed gravitational lensing correspond to a converging or diverging lens? Explain briefly.

Exercise:

Problem:

Suppose you measure the red shifts of all the images produced by gravitational lensing, such as in [\[link\]](#). You find that the central image has a red shift less than the outer images, and those all have the same red shift. Discuss how this not only shows that the images are of the same object, but also implies that the red shift is not affected by taking different paths through space. Does it imply that cosmological red shifts are not caused by traveling through space (light getting tired, perhaps)?

Exercise:**Problem:**

What are gravitational waves, and have they yet been observed either directly or indirectly?

Exercise:**Problem:**

Is the event horizon of a black hole the actual physical surface of the object?

Exercise:**Problem:**

Suppose black holes radiate their mass away and the lifetime of a black hole created by a supernova is about 10^{67} years. How does this lifetime compare with the accepted age of the universe? Is it surprising that we do not observe the predicted characteristic radiation?

Problems & Exercises**Exercise:**

Problem:

What is the Schwarzschild radius of a black hole that has a mass eight times that of our Sun? Note that stars must be more massive than the Sun to form black holes as a result of a supernova.

Solution:

23.6 km

Exercise:**Problem:**

Black holes with masses smaller than those formed in supernovas may have been created in the Big Bang. Calculate the radius of one that has a mass equal to the Earth's.

Exercise:**Problem:**

Supermassive black holes are thought to exist at the center of many galaxies.

- (a) What is the radius of such an object if it has a mass of 10^9 Suns?
 - (b) What is this radius in light years?
-

Solution:

(a) 2.95×10^{12} m

(b) 3.12×10^{-4} ly

Exercise:**Problem: Construct Your Own Problem**

Consider a supermassive black hole near the center of a galaxy. Calculate the radius of such an object based on its mass. You must consider how much mass is reasonable for these large objects, and which is now nearly directly observed. (Information on black holes posted on the Web by NASA and other agencies is reliable, for example.)

Glossary

black holes

objects having such large gravitational fields that things can fall in, but nothing, not even light, can escape

general relativity

Einstein's theory that describes all types of relative motion including accelerated motion and the effects of gravity

gravitational waves

mass-created distortions in space that propagate at the speed of light and that are predicted by general relativity

escape velocity

takeoff velocity when kinetic energy just cancels gravitational potential energy

event horizon

the distance from the object at which the escape velocity is exactly the speed of light

neutron stars

literally a star composed of neutrons

Schwarzschild radius

the radius of the event horizon

thought experiment

mental analysis of certain carefully and clearly defined situations to develop an idea

quasars

the moderately distant galaxies that emit as much or more energy than a normal galaxy

Quantum gravity

the theory that deals with particle exchange of gravitons as the mechanism for the force

Superstrings

- Define Superstring theory.
- Explain the relationship between Superstring theory and the Big Bang.

Introduced earlier in [GUTS: The Unification of Forces](#) **Superstring theory** is an attempt to unify gravity with the other three forces and, thus, must contain quantum gravity. The main tenet of Superstring theory is that fundamental particles, including the graviton that carries the gravitational force, act like one-dimensional vibrating strings. Since gravity affects the time and space in which all else exists, Superstring theory is an attempt at a Theory of Everything (TOE). Each independent quantum number is thought of as a separate dimension in some super space (analogous to the fact that the familiar dimensions of space are independent of one another) and is represented by a different type of Superstring. As the universe evolved after the Big Bang and forces became distinct (spontaneous symmetry breaking), some of the dimensions of superspace are imagined to have curled up and become unnoticed.

Forces are expected to be unified only at extremely high energies and at particle separations on the order of 10^{-35} m. This could mean that Superstrings must have dimensions or wavelengths of this size or smaller. Just as quantum gravity may imply that there are no time intervals shorter than some finite value, it also implies that there may be no sizes smaller than some tiny but finite value. That may be about 10^{-35} m. If so, and if Superstring theory can explain all it strives to, then the structures of Superstrings are at the lower limit of the smallest possible size and can have no further substructure. This would be the ultimate answer to the question the ancient Greeks considered. There is a finite lower limit to space.

Not only is Superstring theory in its infancy, it deals with dimensions about 17 orders of magnitude smaller than the 10^{-18} m details that we have been able to observe directly. It is thus relatively unconstrained by experiment, and there are a host of theoretical possibilities to choose from. This has led theorists to make choices subjectively (as always) on what is the most elegant theory, with less hope than usual that experiment will guide them. It has also led to speculation of alternate universes, with their Big Bangs

creating each new universe with a random set of rules. These speculations may not be tested even in principle, since an alternate universe is by definition unattainable. It is something like exploring a self-consistent field of mathematics, with its axioms and rules of logic that are not consistent with nature. Such endeavors have often given insight to mathematicians and scientists alike and occasionally have been directly related to the description of new discoveries.

Section Summary

- Superstring theory holds that fundamental particles are one-dimensional vibrations analogous to those on strings and is an attempt at a theory of quantum gravity.

Problems & Exercises

Exercise:

Problem:

The characteristic length of entities in Superstring theory is approximately 10^{-35} m.

- (a) Find the energy in GeV of a photon of this wavelength.
- (b) Compare this with the average particle energy of 10^{19} GeV needed for unification of forces.

Solution:

- (a) 1×10^{20}
- (b) 10 times greater

Glossary

Superstring theory

a theory to unify gravity with the other three forces in which the fundamental particles are considered to act like one-dimensional vibrating strings

Dark Matter and Closure

- Discuss the existence of dark matter.
- Explain neutrino oscillations and their consequences.

One of the most exciting problems in physics today is the fact that there is far more matter in the universe than we can see. The motion of stars in galaxies and the motion of galaxies in clusters imply that there is about 10 times as much mass as in the luminous objects we can see. The indirectly observed non-luminous matter is called **dark matter**. Why is dark matter a problem? For one thing, we do not know what it is. It may well be 90% of all matter in the universe, yet there is a possibility that it is of a completely unknown form—a stunning discovery if verified. Dark matter has implications for particle physics. It may be possible that neutrinos actually have small masses or that there are completely unknown types of particles. Dark matter also has implications for cosmology, since there may be enough dark matter to stop the expansion of the universe. That is another problem related to dark matter—we do not know how much there is. We keep finding evidence for more matter in the universe, and we have an idea of how much it would take to eventually stop the expansion of the universe, but whether there is enough is still unknown.

Evidence

The first clues that there is more matter than meets the eye came from the Swiss-born American astronomer Fritz Zwicky in the 1930s; some initial work was also done by the American astronomer Vera Rubin. Zwicky measured the velocities of stars orbiting the galaxy, using the relativistic Doppler shift of their spectra (see [\[link\]](#)(a)). He found that velocity varied with distance from the center of the galaxy, as graphed in [\[link\]](#)(b). If the mass of the galaxy was concentrated in its center, as are its luminous stars, the velocities should decrease as the square root of the distance from the center. Instead, the velocity curve is almost flat, implying that there is a tremendous amount of matter in the galactic halo. Although not immediately recognized for its significance, such measurements have now been made for many galaxies, with similar results. Further, studies of galactic clusters have also indicated that galaxies have a mass distribution

greater than that obtained from their brightness (proportional to the number of stars), which also extends into large halos surrounding the luminous parts of galaxies. Observations of other EM wavelengths, such as radio waves and X rays, have similarly confirmed the existence of dark matter. Take, for example, X rays in the relatively dark space between galaxies, which indicates the presence of previously unobserved hot, ionized gas (see [\[link\]](#) (c)).

Theoretical Yearnings for Closure

Is the universe open or closed? That is, will the universe expand forever or will it stop, perhaps to contract? This, until recently, was a question of whether there is enough gravitation to stop the expansion of the universe. In the past few years, it has become a question of the combination of gravitation and what is called the **cosmological constant**. The cosmological constant was invented by Einstein to prohibit the expansion or contraction of the universe. At the time he developed general relativity, Einstein considered that an illogical possibility. The cosmological constant was discarded after Hubble discovered the expansion, but has been re-invoked in recent years.

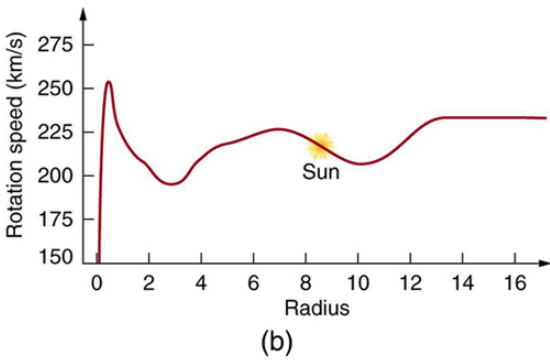
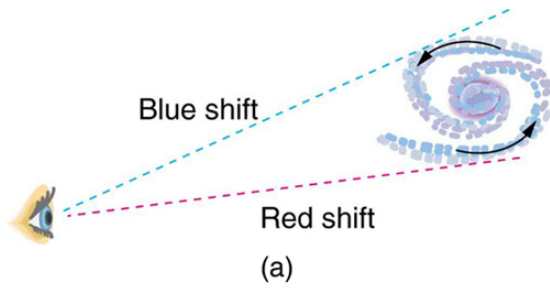
Gravitational attraction between galaxies is slowing the expansion of the universe, but the amount of slowing down is not known directly. In fact, the cosmological constant can counteract gravity's effect. As recent measurements indicate, the universe is expanding *faster* now than in the past—perhaps a “modern inflationary era” in which the dark energy is thought to be causing the expansion of the present-day universe to accelerate. If the expansion rate were affected by gravity alone, we should be able to see that the expansion rate between distant galaxies was once greater than it is now. However, measurements show it was *less* than now. We can, however, calculate the amount of slowing based on the average density of matter we observe directly. Here we have a definite answer—there is far less visible matter than needed to stop expansion. The **critical density** ρ_c is defined to be the density needed to just halt universal expansion in a universe with no cosmological constant. It is estimated to be about

Equation:

$$\rho_c \approx 10^{-26} \text{ kg/m}^3.$$

However, this estimate of ρ_c is only good to about a factor of two, due to uncertainties in the expansion rate of the universe. The critical density is equivalent to an average of only a few nucleons per cubic meter, remarkably small and indicative of how truly empty intergalactic space is. Luminous matter seems to account for roughly 0.5% to 2% of the critical density, far less than that needed for closure. Taking into account the amount of dark matter we detect indirectly and all other types of indirectly observed normal matter, there is only 10% to 40% of what is needed for closure. If we are able to refine the measurements of expansion rates now and in the past, we will have our answer regarding the curvature of space and we will determine a value for the cosmological constant to justify this observation. Finally, the most recent measurements of the CMBR have implications for the cosmological constant, so it is not simply a device concocted for a single purpose.

After the recent experimental discovery of the cosmological constant, most researchers feel that the universe should be just barely open. Since matter can be thought to curve the space around it, we call an open universe **negatively curved**. This means that you can in principle travel an unlimited distance in any direction. A universe that is closed is called **positively curved**. This means that if you travel far enough in any direction, you will return to your starting point, analogous to circumnavigating the Earth. In between these two is a **flat (zero curvature) universe**. The recent discovery of the cosmological constant has shown the universe is very close to flat, and will expand forever. Why do theorists feel the universe is flat? Flatness is a part of the inflationary scenario that helps explain the flatness of the microwave background. In fact, since general relativity implies that matter creates the space in which it exists, there is a special symmetry to a flat universe.



Evidence for dark matter: (a)

We can measure the velocities of stars relative to their galaxies by observing the Doppler shift in emitted light, usually using the hydrogen spectrum. These measurements indicate the

rotation of a spiral galaxy.
(b) A graph of velocity versus distance from the galactic center shows that the velocity does not decrease as it would if the matter were concentrated in luminous stars. The flatness of the curve implies a massive galactic halo of dark matter extending beyond the visible stars. (c) This is a computer-generated image of X rays from a galactic cluster. The X rays indicate the presence of otherwise unseen hot clouds of ionized gas in the regions of space previously considered more empty. (credit: NASA, ESA, CXC, M. Bradac (University of California, Santa Barbara), and S. Allen (Stanford University))

What Is the Dark Matter We See Indirectly?

There is no doubt that dark matter exists, but its form and the amount in existence are two facts that are still being studied vigorously. As always, we seek to explain new observations in terms of known principles. However, as more discoveries are made, it is becoming more and more difficult to explain dark matter as a known type of matter.

One of the possibilities for normal matter is being explored using the Hubble Space Telescope and employing the lensing effect of gravity on

light (see [\[link\]](#)). Stars glow because of nuclear fusion in them, but planets are visible primarily by reflected light. Jupiter, for example, is too small to ignite fusion in its core and become a star, but we can see sunlight reflected from it, since we are relatively close. If Jupiter orbited another star, we would not be able to see it directly. The question is open as to how many planets or other bodies smaller than about 1/1000 the mass of the Sun are there. If such bodies pass between us and a star, they will not block the star's light, being too small, but they will form a gravitational lens, as discussed in [General Relativity and Quantum Gravity](#).

In a process called **microlensing**, light from the star is focused and the star appears to brighten in a characteristic manner. Searches for dark matter in this form are particularly interested in galactic halos because of the huge amount of mass that seems to be there. Such microlensing objects are thus called **massive compact halo objects**, or **MACHOs**. To date, a few MACHOs have been observed, but not predominantly in galactic halos, nor in the numbers needed to explain dark matter.

MACHOs are among the most conventional of unseen objects proposed to explain dark matter. Others being actively pursued are red dwarfs, which are small dim stars, but too few have been seen so far, even with the Hubble Telescope, to be of significance. Old remnants of stars called white dwarfs are also under consideration, since they contain about a solar mass, but are small as the Earth and may dim to the point that we ordinarily do not observe them. While white dwarfs are known, old dim ones are not. Yet another possibility is the existence of large numbers of smaller than stellar mass black holes left from the Big Bang—here evidence is entirely absent.

There is a very real possibility that dark matter is composed of the known neutrinos, which may have small, but finite, masses. As discussed earlier, neutrinos are thought to be massless, but we only have upper limits on their masses, rather than knowing they are exactly zero. So far, these upper limits come from difficult measurements of total energy emitted in the decays and reactions in which neutrinos are involved. There is an amusing possibility of proving that neutrinos have mass in a completely different way.

We have noted in [Particles, Patterns, and Conservation Laws](#) that there are three flavors of neutrinos (ν_e , ν_μ , and ν_τ) and that the weak interaction could change quark flavor. It should also change neutrino flavor—that is, any type of neutrino could change spontaneously into any other, a process called **neutrino oscillations**. However, this can occur only if neutrinos have a mass. Why? Crudely, because if neutrinos are massless, they must travel at the speed of light and time will not pass for them, so that they cannot change without an interaction. In 1999, results began to be published containing convincing evidence that neutrino oscillations do occur. Using the Super-Kamiokande detector in Japan, the oscillations have been observed and are being verified and further explored at present at the same facility and others.

Neutrino oscillations may also explain the low number of observed solar neutrinos. Detectors for observing solar neutrinos are specifically designed to detect electron neutrinos ν_e produced in huge numbers by fusion in the Sun. A large fraction of electron neutrinos ν_e may be changing flavor to muon neutrinos ν_μ on their way out of the Sun, possibly enhanced by specific interactions, reducing the flux of electron neutrinos to observed levels. There is also a discrepancy in observations of neutrinos produced in cosmic ray showers. While these showers of radiation produced by extremely energetic cosmic rays should contain twice as many ν_μ s as ν_e s, their numbers are nearly equal. This may be explained by neutrino oscillations from muon flavor to electron flavor. Massive neutrinos are a particularly appealing possibility for explaining dark matter, since their existence is consistent with a large body of known information and explains more than dark matter. The question is not settled at this writing.

The most radical proposal to explain dark matter is that it consists of previously unknown leptons (sometimes obtusely referred to as non-baryonic matter). These are called **weakly interacting massive particles**, or **WIMPs**, and would also be chargeless, thus interacting negligibly with normal matter, except through gravitation. One proposed group of WIMPs would have masses several orders of magnitude greater than nucleons and are sometimes called **neutralinos**. Others are called **axions** and would have masses about 10^{-10} that of an electron mass. Both neutralinos and axions would be gravitationally attached to galaxies, but because they are

chargeless and only feel the weak force, they would be in a halo rather than interact and coalesce into spirals, and so on, like normal matter (see [\[link\]](#)).



The Hubble Space Telescope is producing exciting data with its corrected optics and with the absence of atmospheric distortion. It has observed some MACHOs, disks of material around stars thought to precede planet formation, black hole candidates, and collisions of comets with Jupiter. (credit: NASA (crew of STS-125))



Dark matter may shepherd normal matter gravitationally in space, as this stream moves the leaves. Dark matter may be invisible and even move through the normal matter, as neutrinos penetrate us without small-scale effect. (credit: Shinichi Sugiyama)

Some particle theorists have built WIMPs into their unified force theories and into the inflationary scenario of the evolution of the universe so popular today. These particles would have been produced in just the correct numbers to make the universe flat, shortly after the Big Bang. The proposal is radical in the sense that it invokes entirely new forms of matter, in fact *two* entirely new forms, in order to explain dark matter and other phenomena. WIMPs have the extra burden of automatically being very difficult to observe directly. This is somewhat analogous to quark confinement, which guarantees that quarks are there, but they can never be seen directly. One of the primary goals of the LHC at CERN, however, is to produce and detect WIMPs. At any rate, before WIMPs are accepted as the best explanation, all other possibilities utilizing known phenomena will have to be shown inferior. Should that occur, we will be in the unanticipated position of admitting that, to date, all we know is only 10% of what exists.

A far cry from the days when people firmly believed themselves to be not only the center of the universe, but also the reason for its existence.

Section Summary

- Dark matter is non-luminous matter detected in and around galaxies and galactic clusters.
- It may be 10 times the mass of the luminous matter in the universe, and its amount may determine whether the universe is open or closed (expands forever or eventually stops).
- The determining factor is the critical density of the universe and the cosmological constant, a theoretical construct intimately related to the expansion and closure of the universe.
- The critical density ρ_c is the density needed to just halt universal expansion. It is estimated to be approximately 10^{-26} kg/m^3 .
- An open universe is negatively curved, a closed universe is positively curved, whereas a universe with exactly the critical density is flat.
- Dark matter's composition is a major mystery, but it may be due to the suspected mass of neutrinos or a completely unknown type of leptonic matter.
- If neutrinos have mass, they will change families, a process known as neutrino oscillations, for which there is growing evidence.

Conceptual Questions

Exercise:

Problem:

Discuss the possibility that star velocities at the edges of galaxies being greater than expected is due to unknown properties of gravity rather than to the existence of dark matter. Would this mean, for example, that gravity is greater or smaller than expected at large distances? Are there other tests that could be made of gravity at large distances, such as observing the motions of neighboring galaxies?

Exercise:

Problem:

How does relativistic time dilation prohibit neutrino oscillations if they are massless?

Exercise:**Problem:**

If neutrino oscillations do occur, will they violate conservation of the various lepton family numbers (L_e , L_μ , and L_τ)? Will neutrino oscillations violate conservation of the total number of leptons?

Exercise:**Problem:**

Lacking direct evidence of WIMPs as dark matter, why must we eliminate all other possible explanations based on the known forms of matter before we invoke their existence?

Problems Exercises**Exercise:****Problem:**

If the dark matter in the Milky Way were composed entirely of MACHOs (evidence shows it is not), approximately how many would there have to be? Assume the average mass of a MACHO is 1/1000 that of the Sun, and that dark matter has a mass 10 times that of the luminous Milky Way galaxy with its 10^{11} stars of average mass 1.5 times the Sun's mass.

Solution:**Equation:**

$$1.5 \times 10^{15}$$

Exercise:**Problem:**

The critical mass density needed to just halt the expansion of the universe is approximately 10^{-26} kg/m^3 .

(a) Convert this to $\text{eV}/c^2 \cdot \text{m}^3$.

(b) Find the number of neutrinos per cubic meter needed to close the universe if their average mass is $7 \text{ eV}/c^2$ and they have negligible kinetic energies.

Exercise:**Problem:**

Assume the average density of the universe is 0.1 of the critical density needed for closure. What is the average number of protons per cubic meter, assuming the universe is composed mostly of hydrogen?

Solution:**Equation:**

$$0.6 \text{ m}^{-3}$$

Exercise:**Problem:**

To get an idea of how empty deep space is on the average, perform the following calculations:

(a) Find the volume our Sun would occupy if it had an average density equal to the critical density of 10^{-26} kg/m^3 thought necessary to halt the expansion of the universe.

(b) Find the radius of a sphere of this volume in light years.

(c) What would this radius be if the density were that of luminous matter, which is approximately 5% that of the critical density?

(d) Compare the radius found in part (c) with the 4-ly average separation of stars in the arms of the Milky Way.

Glossary

axions

a type of WIMPs having masses about 10^{-10} of an electron mass

cosmological constant

a theoretical construct intimately related to the expansion and closure of the universe

critical density

the density of matter needed to just halt universal expansion

dark matter

indirectly observed non-luminous matter

flat (zero curvature) universe

a universe that is infinite but not curved

microlensing

a process in which light from a distant star is focused and the star appears to brighten in a characteristic manner, when a small body (smaller than about 1/1000 the mass of the Sun) passes between us and the star

MACHOs

massive compact halo objects; microlensing objects of huge mass

neutrino oscillations

a process in which any type of neutrino could change spontaneously into any other

neutralinos

a type of WIMPs having masses several orders of magnitude greater than nucleon masses

negatively curved

an open universe that expands forever

positively curved

a universe that is closed and eventually contracts

WIMPs

weakly interacting massive particles; chargeless leptons (non-baryonic matter) interacting negligibly with normal matter

Complexity and Chaos

- Explain complex systems.
- Discuss chaotic behavior of different systems.

Much of what impresses us about physics is related to the underlying connections and basic simplicity of the laws we have discovered. The language of physics is precise and well defined because many basic systems we study are simple enough that we can perform controlled experiments and discover unambiguous relationships. Our most spectacular successes, such as the prediction of previously unobserved particles, come from the simple underlying patterns we have been able to recognize. But there are systems of interest to physicists that are inherently complex. The simple laws of physics apply, of course, but complex systems may reveal patterns that simple systems do not. The emerging field of **complexity** is devoted to the study of complex systems, including those outside the traditional bounds of physics. Of particular interest is the ability of complex systems to adapt and evolve.

What are some examples of complex adaptive systems? One is the primordial ocean. When the oceans first formed, they were a random mix of elements and compounds that obeyed the laws of physics and chemistry. In a relatively short geological time (about 500 million years), life had emerged. Laboratory simulations indicate that the emergence of life was far too fast to have come from random combinations of compounds, even if driven by lightning and heat. There must be an underlying ability of the complex system to organize itself, resulting in the self-replication we recognize as life. Living entities, even at the unicellular level, are highly organized and systematic. Systems of living organisms are themselves complex adaptive systems. The grandest of these evolved into the biological system we have today, leaving traces in the geological record of steps taken along the way.

Complexity as a discipline examines complex systems, how they adapt and evolve, looking for similarities with other complex adaptive systems. Can, for example, parallels be drawn between biological evolution and the evolution of *economic systems*? Economic systems do emerge quickly, they show tendencies for self-organization, they are complex (in the number and

types of transactions), and they adapt and evolve. Biological systems do all the same types of things. There are other examples of complex adaptive systems being studied for fundamental similarities. *Cultures* show signs of adaptation and evolution. The comparison of different cultural evolutions may bear fruit as well as comparisons to biological evolution. *Science* also is a complex system of human interactions, like culture and economics, that adapts to new information and political pressure, and evolves, usually becoming more organized rather than less. Those who study *creative thinking* also see parallels with complex systems. Humans sometimes organize almost random pieces of information, often subconsciously while doing other things, and come up with brilliant creative insights. The development of *language* is another complex adaptive system that may show similar tendencies. *Artificial intelligence* is an overt attempt to devise an adaptive system that will self-organize and evolve in the same manner as an intelligent living being learns. These are a few of the broad range of topics being studied by those who investigate complexity. There are now institutes, journals, and meetings, as well as popularizations of the emerging topic of complexity.

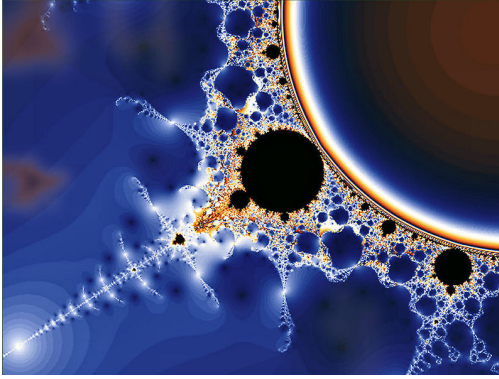
In traditional physics, the discipline of complexity may yield insights in certain areas. Thermodynamics treats systems on the average, while statistical mechanics deals in some detail with complex systems of atoms and molecules in random thermal motion. Yet there is organization, adaptation, and evolution in those complex systems. Non-equilibrium phenomena, such as heat transfer and phase changes, are characteristically complex in detail, and new approaches to them may evolve from complexity as a discipline. Crystal growth is another example of self-organization spontaneously emerging in a complex system. Alloys are also inherently complex mixtures that show certain simple characteristics implying some self-organization. The organization of iron atoms into magnetic domains as they cool is another. Perhaps insights into these difficult areas will emerge from complexity. But at the minimum, the discipline of complexity is another example of human effort to understand and organize the universe around us, partly rooted in the discipline of physics.

A predecessor to complexity is the topic of chaos, which has been widely publicized and has become a discipline of its own. It is also based partly in physics and treats broad classes of phenomena from many disciplines.

Chaos is a word used to describe systems whose outcomes are extremely sensitive to initial conditions. The orbit of the planet Pluto, for example, may be chaotic in that it can change tremendously due to small interactions with other planets. This makes its long-term behavior impossible to predict with precision, just as we cannot tell precisely where a decaying Earth satellite will land or how many pieces it will break into. But the discipline of chaos has found ways to deal with such systems and has been applied to apparently unrelated systems. For example, the heartbeat of people with certain types of potentially lethal arrhythmias seems to be chaotic, and this knowledge may allow more sophisticated monitoring and recognition of the need for intervention.

Chaos is related to complexity. Some chaotic systems are also inherently complex; for example, vortices in a fluid as opposed to a double pendulum. Both are chaotic and not predictable in the same sense as other systems. But there can be organization in chaos and it can also be quantified. Examples of chaotic systems are beautiful fractal patterns such as in [\[link\]](#). Some chaotic systems exhibit self-organization, a type of stable chaos. The orbits of the planets in our solar system, for example, may be chaotic (we are not certain yet). But they are definitely organized and systematic, with a simple formula describing the orbital radii of the first eight planets *and* the asteroid belt. Large-scale vortices in Jupiter's atmosphere are chaotic, but the Great Red Spot is a stable self-organization of rotational energy. (See [\[link\]](#).) The Great Red Spot has been in existence for at least 400 years and is a complex self-adaptive system.

The emerging field of complexity, like the now almost traditional field of chaos, is partly rooted in physics. Both attempt to see similar systematics in a very broad range of phenomena and, hence, generate a better understanding of them. Time will tell what impact these fields have on more traditional areas of physics as well as on the other disciplines they relate to.



This image is related to the Mandelbrot set, a complex mathematical form that is chaotic. The patterns are infinitely fine as you look closer and closer, and they indicate order in the presence of chaos. (credit: Gilberto Santa Rosa)



The Great Red Spot on Jupiter is an example of self-organization in a complex and chaotic system. Smaller vortices in Jupiter's atmosphere

behave chaotically, but
the triple-Earth-size spot
is self-organized and
stable for at least
hundreds of years. (credit:
NASA)

Section Summary

- Complexity is an emerging field, rooted primarily in physics, that considers complex adaptive systems and their evolution, including self-organization.
- Complexity has applications in physics and many other disciplines, such as biological evolution.
- Chaos is a field that studies systems whose properties depend extremely sensitively on some variables and whose evolution is impossible to predict.
- Chaotic systems may be simple or complex.
- Studies of chaos have led to methods for understanding and predicting certain chaotic behaviors.

Conceptual Questions

Exercise:

Problem:

Must a complex system be adaptive to be of interest in the field of complexity? Give an example to support your answer.

Exercise:

Problem: State a necessary condition for a system to be chaotic.

Glossary

complexity

an emerging field devoted to the study of complex systems

chaos

word used to describe systems the outcomes of which are extremely sensitive to initial conditions

High-temperature Superconductors

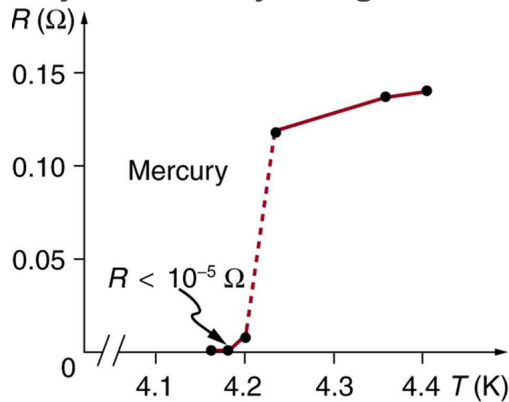
- Identify superconductors and their uses.
- Discuss the need for a high- T_c superconductor.

Superconductors are materials with a resistivity of zero. They are familiar to the general public because of their practical applications and have been mentioned at a number of points in the text. Because the resistance of a piece of superconductor is zero, there are no heat losses for currents through them; they are used in magnets needing high currents, such as in MRI machines, and could cut energy losses in power transmission. But most superconductors must be cooled to temperatures only a few kelvin above absolute zero, a costly procedure limiting their practical applications. In the past decade, tremendous advances have been made in producing materials that become superconductors at relatively high temperatures. There is hope that room temperature superconductors may someday be manufactured.

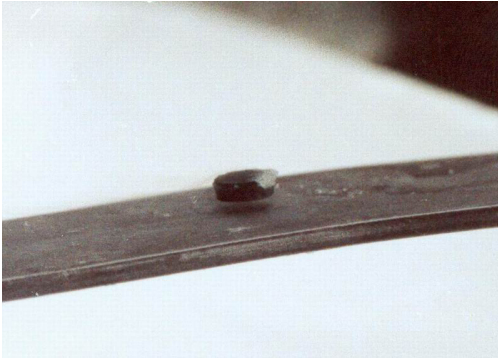
Superconductivity was discovered accidentally in 1911 by the Dutch physicist H. Kamerlingh Onnes (1853–1926) when he used liquid helium to cool mercury. Onnes had been the first person to liquefy helium a few years earlier and was surprised to observe the resistivity of a mediocre conductor like mercury drop to zero at a temperature of 4.2 K. We define the temperature at which and below which a material becomes a superconductor to be its **critical temperature**, denoted by T_c . (See [\[link\]](#).) Progress in understanding how and why a material became a superconductor was relatively slow, with the first workable theory coming in 1957. Certain other elements were also found to become superconductors, but all had T_c s less than 10 K, which are expensive to maintain. Although Onnes received a Nobel prize in 1913, it was primarily for his work with liquid helium.

In 1986, a breakthrough was announced—a ceramic compound was found to have an unprecedented T_c of 35 K. It looked as if much higher critical temperatures could be possible, and by early 1988 another ceramic (this of thallium, calcium, barium, copper, and oxygen) had been found to have $T_c = 125$ K (see [\[link\]](#).) The economic potential of perfect conductors saving electric energy is immense for T_c s above 77 K, since that is the temperature of liquid nitrogen. Although liquid helium has a boiling point

of 4 K and can be used to make materials superconducting, it costs about \$5 per liter. Liquid nitrogen boils at 77 K, but only costs about \$0.30 per liter. There was general euphoria at the discovery of these complex ceramic superconductors, but this soon subsided with the sobering difficulty of forming them into usable wires. The first commercial use of a high temperature superconductor is in an electronic filter for cellular phones. High-temperature superconductors are used in experimental apparatus, and they are actively being researched, particularly in thin film applications.



A graph of resistivity versus temperature for a superconductor shows a sharp transition to zero at the critical temperature T_c . High temperature superconductors have verifiable T_c s greater than 125 K, well above the easily achieved 77-K temperature of liquid nitrogen.

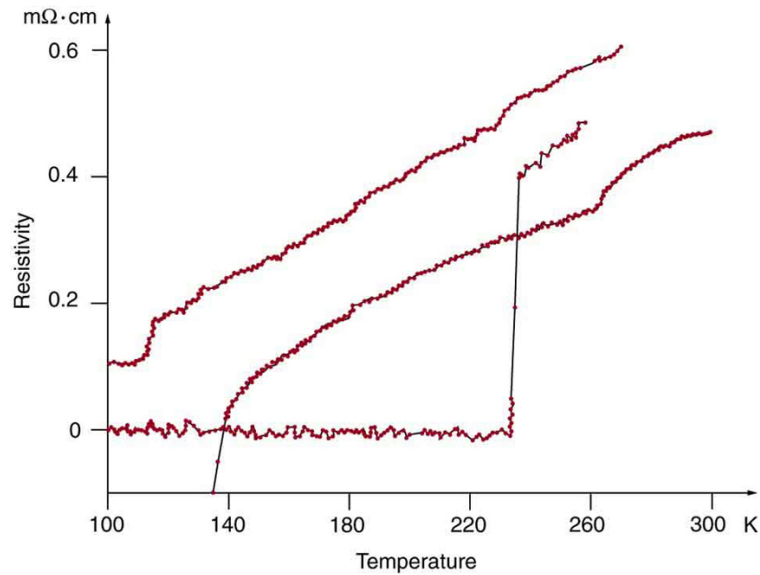


One characteristic of a superconductor is that it excludes magnetic flux and, thus, repels other magnets. The small magnet levitated above a high-temperature superconductor, which is cooled by liquid nitrogen, gives evidence that the material is superconducting. When the material warms and becomes conducting, magnetic flux can penetrate it, and the magnet will rest upon it.
(credit: Saperaud)

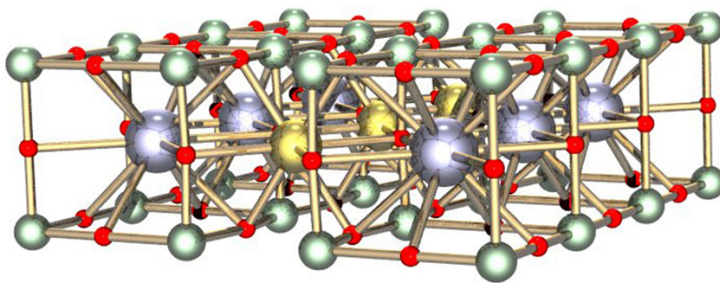
The search is on for even higher T_c superconductors, many of complex and exotic copper oxide ceramics, sometimes including strontium, mercury, or yttrium as well as barium, calcium, and other elements. Room temperature (about 293 K) would be ideal, but any temperature close to room temperature is relatively cheap to produce and maintain. There are persistent reports of T_c s over 200 K and some in the vicinity of 270 K. Unfortunately, these observations are not routinely reproducible, with

samples losing their superconducting nature once heated and re-cooled (cycled) a few times (see [\[link\]](#).) They are now called USOs or unidentified superconducting objects, out of frustration and the refusal of some samples to show high T_c even though produced in the same manner as others. Reproducibility is crucial to discovery, and researchers are justifiably reluctant to claim the breakthrough they all seek. Time will tell whether USOs are real or an experimental quirk.

The theory of ordinary superconductors is difficult, involving quantum effects for widely separated electrons traveling through a material. Electrons couple in a manner that allows them to get through the material without losing energy to it, making it a superconductor. High- T_c superconductors are more difficult to understand theoretically, but theorists seem to be closing in on a workable theory. The difficulty of understanding how electrons can sneak through materials without losing energy in collisions is even greater at higher temperatures, where vibrating atoms should get in the way. Discoverers of high T_c may feel something analogous to what a politician once said upon an unexpected election victory—“I wonder what we did right?”



(a)



(b)

(a) This graph, adapted from an article in *Physics Today*, shows the behavior of a single sample of a high-temperature superconductor in three different trials. In one case the sample exhibited a T_c of about 230 K, whereas in the others it did not become superconducting at all. The lack of reproducibility is typical of forefront experiments and prohibits definitive conclusions. (b) This colorful diagram shows the complex but systematic nature of the lattice structure of a high-temperature superconducting

ceramic. (credit: en:Cadmium,
Wikimedia Commons)

Section Summary

- High-temperature superconductors are materials that become superconducting at temperatures well above a few kelvin.
- The critical temperature T_c is the temperature below which a material is superconducting.
- Some high-temperature superconductors have verified T_c s above 125 K, and there are reports of T_c s as high as 250 K.

Conceptual Questions

Exercise:

Problem:

What is critical temperature T_c ? Do all materials have a critical temperature? Explain why or why not.

Exercise:

Problem:

Explain how good thermal contact with liquid nitrogen can keep objects at a temperature of 77 K (liquid nitrogen's boiling point at atmospheric pressure).

Exercise:

Problem:

Not only is liquid nitrogen a cheaper coolant than liquid helium, its boiling point is higher (77 K vs. 4.2 K). How does higher temperature help lower the cost of cooling a material? Explain in terms of the rate of heat transfer being related to the temperature difference between the sample and its surroundings.

Problem Exercises

Exercise:

Problem:

A section of superconducting wire carries a current of 100 A and requires 1.00 L of liquid nitrogen per hour to keep it below its critical temperature. For it to be economically advantageous to use a superconducting wire, the cost of cooling the wire must be less than the cost of energy lost to heat in the wire. Assume that the cost of liquid nitrogen is \$0.30 per liter, and that electric energy costs \$0.10 per kW·h. What is the resistance of a normal wire that costs as much in wasted electric energy as the cost of liquid nitrogen for the superconductor?

Solution:

Equation:

$$0.30 \, \Omega$$

Glossary

Superconductors

materials with resistivity of zero

critical temperature

the temperature at which and below which a material becomes a superconductor

Some Questions We Know to Ask

- Identify sample questions to be asked on the largest scales.
- Identify sample questions to be asked on the intermediate scale.
- Identify sample questions to be asked on the smallest scales.

Throughout the text we have noted how essential it is to be curious and to ask questions in order to first understand what is known, and then to go a little farther. Some questions may go unanswered for centuries; others may not have answers, but some bear delicious fruit. Part of discovery is knowing which questions to ask. You have to know something before you can even phrase a decent question. As you may have noticed, the mere act of asking a question can give you the answer. The following questions are a sample of those physicists now know to ask and are representative of the forefronts of physics. Although these questions are important, they will be replaced by others if answers are found to them. The fun continues.

On the Largest Scale

1. *Is the universe open or closed?* Theorists would like it to be just barely closed and evidence is building toward that conclusion. Recent measurements in the expansion rate of the universe and in CMBR support a flat universe. There is a connection to small-scale physics in the type and number of particles that may contribute to closing the universe.
2. *What is dark matter?* It is definitely there, but we really do not know what it is. Conventional possibilities are being ruled out, but one of them still may explain it. The answer could reveal whole new realms of physics and the disturbing possibility that most of what is out there is unknown to us, a completely different form of matter.
3. *How do galaxies form?* They exist since very early in the evolution of the universe and it remains difficult to understand how they evolved so quickly. The recent finer measurements of fluctuations in the CMBR may yet allow us to explain galaxy formation.
4. *What is the nature of various-mass black holes?* Only recently have we become confident that many black hole candidates cannot be explained by other, less exotic possibilities. But we still do not know much about

how they form, what their role in the history of galactic evolution has been, and the nature of space in their vicinity. However, so many black holes are now known that correlations between black hole mass and galactic nuclei characteristics are being studied.

5. *What is the mechanism for the energy output of quasars?* These distant and extraordinarily energetic objects now seem to be early stages of galactic evolution with a supermassive black-hole-devouring material. Connections are now being made with galaxies having energetic cores, and there is evidence consistent with less consuming, supermassive black holes at the center of older galaxies. New instruments are allowing us to see deeper into our own galaxy for evidence of our own massive black hole.
6. *Where do the γ bursts come from?* We see bursts of γ rays coming from all directions in space, indicating the sources are very distant objects rather than something associated with our own galaxy. Some γ bursts finally are being correlated with known sources so that the possibility they may originate in binary neutron star interactions or black holes eating a companion neutron star can be explored.

On the Intermediate Scale

1. *How do phase transitions take place on the microscopic scale?* We know a lot about phase transitions, such as water freezing, but the details of how they occur molecule by molecule are not well understood. Similar questions about specific heat a century ago led to early quantum mechanics. It is also an example of a complex adaptive system that may yield insights into other self-organizing systems.
2. *Is there a way to deal with nonlinear phenomena that reveals underlying connections?* Nonlinear phenomena lack a direct or linear proportionality that makes analysis and understanding a little easier. There are implications for nonlinear optics and broader topics such as chaos.
3. *How do high- T_c superconductors become resistanceless at such high temperatures?* Understanding how they work may help make them more practical or may result in surprises as unexpected as the discovery of superconductivity itself.

4. *There are magnetic effects in materials we do not understand—how do they work?* Although beyond the scope of this text, there is a great deal to learn in condensed matter physics (the physics of solids and liquids). We may find surprises analogous to lasing, the quantum Hall effect, and the quantization of magnetic flux. Complexity may play a role here, too.

On the Smallest Scale

1. *Are quarks and leptons fundamental, or do they have a substructure?* The higher energy accelerators that are just completed or being constructed may supply some answers, but there will also be input from cosmology and other systematics.
2. *Why do leptons have integral charge while quarks have fractional charge?* If both are fundamental and analogous as thought, this question deserves an answer. It is obviously related to the previous question.
3. *Why are there three families of quarks and leptons?* First, does this imply some relationship? Second, why three and only three families?
4. *Are all forces truly equal (unified) under certain circumstances?* They don't have to be equal just because we want them to be. The answer may have to be indirectly obtained because of the extreme energy at which we think they are unified.
5. *Are there other fundamental forces?* There was a flurry of activity with claims of a fifth and even a sixth force a few years ago. Interest has subsided, since those forces have not been detected consistently. Moreover, the proposed forces have strengths similar to gravity, making them extraordinarily difficult to detect in the presence of stronger forces. But the question remains; and if there are no other forces, we need to ask why only four and why these four.
6. *Is the proton stable?* We have discussed this in some detail, but the question is related to fundamental aspects of the unification of forces. We may never know from experiment that the proton is stable, only that it is very long lived.
7. *Are there magnetic monopoles?* Many particle theories call for very massive individual north- and south-pole particles—magnetic

- monopoles. If they exist, why are they so different in mass and elusiveness from electric charges, and if they do not exist, why not?
8. *Do neutrinos have mass?* Definitive evidence has emerged for neutrinos having mass. The implications are significant, as discussed in this chapter. There are effects on the closure of the universe and on the patterns in particle physics.
 9. *What are the systematic characteristics of high- Z nuclei?* All elements with $Z = 118$ or less (with the exception of 115 and 117) have now been discovered. It has long been conjectured that there may be an island of relative stability near $Z = 114$, and the study of the most recently discovered nuclei will contribute to our understanding of nuclear forces.

These lists of questions are not meant to be complete or consistently important—you can no doubt add to it yourself. There are also important questions in topics not broached in this text, such as certain particle symmetries, that are of current interest to physicists. Hopefully, the point is clear that no matter how much we learn, there always seems to be more to know. Although we are fortunate to have the hard-won wisdom of those who preceded us, we can look forward to new enlightenment, undoubtedly sprinkled with surprise.

Section Summary

- On the largest scale, the questions which can be asked may be about dark matter, dark energy, black holes, quasars, and other aspects of the universe.
- On the intermediate scale, we can query about gravity, phase transitions, nonlinear phenomena, high- T_c superconductors, and magnetic effects on materials.
- On the smallest scale, questions may be about quarks and leptons, fundamental forces, stability of protons, and existence of monopoles.

Conceptual Questions

Exercise:

Problem:

For experimental evidence, particularly of previously unobserved phenomena, to be taken seriously it must be reproducible or of sufficiently high quality that a single observation is meaningful. Supernova 1987A is not reproducible. How do we know observations of it were valid? The fifth force is not broadly accepted. Is this due to lack of reproducibility or poor-quality experiments (or both)? Discuss why forefront experiments are more subject to observational problems than those involving established phenomena.

Exercise:**Problem:**

Discuss whether you think there are limits to what humans can understand about the laws of physics. Support your arguments.

Atomic Masses

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
0	neutron	1	n	1.008 665	β^-	10.37 min
1	Hydrogen	1	^1H	1.007 825	99.985%	
	Deuterium	2	^2H or D	2.014 102	0.015%	
	Tritium	3	^3H or T	3.016 050	β^-	12.33 y
2	Helium	3	^3He	3.016 030	$1.38 \times 10^{-4}\%$	
		4	^4He	4.002 603	$\approx 100\%$	
3	Lithium	6	^6Li	6.015 121	7.5%	
		7	^7Li	7.016 003	92.5%	
4	Beryllium	7	^7Be	7.016 928	EC	53.29 d
		9	^9Be	9.012 182	100%	
5	Boron	10	^{10}B	10.012 937	19.9%	
		11	^{11}B	11.009 305	80.1%	
6	Carbon	11	^{11}C	11.011 432	EC, β^+	
		12	^{12}C	12.000 000	98.90%	
		13	^{13}C	13.003 355	1.10%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		14	^{14}C	14.003 241	β^-	5730 y
7	Nitrogen	13	^{13}N	13.005 738	β^+	9.96 min
		14	^{14}N	14.003 074	99.63%	
		15	^{15}N	15.000 108	0.37%	
8	Oxygen	15	^{15}O	15.003 065	EC, β^+	122 s
		16	^{16}O	15.994 915	99.76%	
		18	^{18}O	17.999 160	0.200%	
9	Fluorine	18	^{18}F	18.000 937	EC, β^+	1.83 h
		19	^{19}F	18.998 403	100%	
10	Neon	20	^{20}Ne	19.992 435	90.51%	
		22	^{22}Ne	21.991 383	9.22%	
11	Sodium	22	^{22}Na	21.994 434	β^+	2.602 y
		23	^{23}Na	22.989 767	100%	
		24	^{24}Na	23.990 961	β^-	14.96 h
12	Magnesium	24	^{24}Mg	23.985 042	78.99%	
13	Aluminum	27	^{27}Al	26.981 539	100%	
14	Silicon	28	^{28}Si	27.976 927	92.23%	2.62h

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		31	^{31}Si	30.975 362	β^-	
15	Phosphorus	31	^{31}P	30.973 762	100%	
		32	^{32}P	31.973 907	β^-	14.28 d
16	Sulfur	32	^{32}S	31.972 070	95.02%	
		35	^{35}S	34.969 031	β^-	87.4 d
17	Chlorine	35	^{35}Cl	34.968 852	75.77%	
		37	^{37}Cl	36.965 903	24.23%	
18	Argon	40	^{40}Ar	39.962 384	99.60%	
19	Potassium	39	^{39}K	38.963 707	93.26%	
		40	^{40}K	39.963 999	0.0117%, EC, β^-	$1.28 \times 10^9 \text{ y}$
20	Calcium	40	^{40}Ca	39.962 591	96.94%	
21	Scandium	45	^{45}Sc	44.955 910	100%	
22	Titanium	48	^{48}Ti	47.947 947	73.8%	
23	Vanadium	51	^{51}V	50.943 962	99.75%	
24	Chromium	52	^{52}Cr	51.940 509	83.79%	
25	Manganese	55	^{55}Mn	54.938 047	100%	
26	Iron	56	^{56}Fe	55.934 939	91.72%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
27	Cobalt	59	^{59}Co	58.933 198	100%	
		60	^{60}Co	59.933 819	β^-	5.271 y
28	Nickel	58	^{58}Ni	57.935 346	68.27%	
		60	^{60}Ni	59.930 788	26.10%	
29	Copper	63	^{63}Cu	62.939 598	69.17%	
		65	^{65}Cu	64.927 793	30.83%	
30	Zinc	64	^{64}Zn	63.929 145	48.6%	
		66	^{66}Zn	65.926 034	27.9%	
31	Gallium	69	^{69}Ga	68.925 580	60.1%	
32	Germanium	72	^{72}Ge	71.922 079	27.4%	
		74	^{74}Ge	73.921 177	36.5%	
33	Arsenic	75	^{75}As	74.921 594	100%	
34	Selenium	80	^{80}Se	79.916 520	49.7%	
35	Bromine	79	^{79}Br	78.918 336	50.69%	
36	Krypton	84	^{84}Kr	83.911 507	57.0%	
37	Rubidium	85	^{85}Rb	84.911 794	72.17%	
38	Strontium	86	^{86}Sr	85.909 267	9.86%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		88	^{88}Sr	87.905 619	82.58%	
		90	^{90}Sr	89.907 738	β^-	28.8 y
39	Yttrium	89	^{89}Y	88.905 849	100%	
		90	^{90}Y	89.907 152	β^-	64.1 h
40	Zirconium	90	^{90}Zr	89.904 703	51.45%	
41	Niobium	93	^{93}Nb	92.906 377	100%	
42	Molybdenum	98	^{98}Mo	97.905 406	24.13%	
43	Technetium	98	^{98}Tc	97.907 215	β^-	$4.2 \times 10^6 \text{ y}$
44	Ruthenium	102	^{102}Ru	101.904 348	31.6%	
45	Rhodium	103	^{103}Rh	102.905 500	100%	
46	Palladium	106	^{106}Pd	105.903 478	27.33%	
47	Silver	107	^{107}Ag	106.905 092	51.84%	
		109	^{109}Ag	108.904 757	48.16%	
48	Cadmium	114	^{114}Cd	113.903 357	28.73%	
49	Indium	115	^{115}In	114.903 880	95.7%, β^-	$4.4 \times 10^{14} \text{ y}$
50	Tin	120	^{120}Sn	119.902 200	32.59%	
51	Antimony	121	^{121}Sb	120.903 821	57.3%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
52	Tellurium	130	^{130}Te	129.906 229	33.8%, β^-	$2.5 \times 10^{21}\text{y}$
53	Iodine	127	^{127}I	126.904 473	100%	
		131	^{131}I	130.906 114	β^-	8.040 d
54	Xenon	132	^{132}Xe	131.904 144	26.9%	
		136	^{136}Xe	135.907 214	8.9%	
55	Cesium	133	^{133}Cs	132.905 429	100%	
		134	^{134}Cs	133.906 696	EC, β^-	2.06 y
56	Barium	137	^{137}Ba	136.905 812	11.23%	
		138	^{138}Ba	137.905 232	71.70%	
57	Lanthanum	139	^{139}La	138.906 346	99.91%	
58	Cerium	140	^{140}Ce	139.905 433	88.48%	
59	Praseodymium	141	^{141}Pr	140.907 647	100%	
60	Neodymium	142	^{142}Nd	141.907 719	27.13%	
61	Promethium	145	^{145}Pm	144.912 743	EC, α	17.7 y
62	Samarium	152	^{152}Sm	151.919 729	26.7%	
63	Europium	153	^{153}Eu	152.921 225	52.2%	
64	Gadolinium	158	^{158}Gd	157.924 099	24.84%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
65	Terbium	159	^{159}Tb	158.925 342	100%	
66	Dysprosium	164	^{164}Dy	163.929 171	28.2%	
67	Holmium	165	^{165}Ho	164.930 319	100%	
68	Erbium	166	^{166}Er	165.930 290	33.6%	
69	Thulium	169	^{169}Tm	168.934 212	100%	
70	Ytterbium	174	^{174}Yb	173.938 859	31.8%	
71	Lutecium	175	^{175}Lu	174.940 770	97.41%	
72	Hafnium	180	^{180}Hf	179.946 545	35.10%	
73	Tantalum	181	^{181}Ta	180.947 992	99.98%	
74	Tungsten	184	^{184}W	183.950 928	30.67%	
75	Rhenium	187	^{187}Re	186.955 744	62.6%, β^-	$4.6 \times 10^{10}\text{y}$
76	Osmium	191	^{191}Os	190.960 920	β^-	15.4 d
		192	^{192}Os	191.961 467	41.0%	
77	Iridium	191	^{191}Ir	190.960 584	37.3%	
		193	^{193}Ir	192.962 917	62.7%	
78	Platinum	195	^{195}Pt	194.964 766	33.8%	
79	Gold	197	^{197}Au	196.966 543	100%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		198	^{198}Au	197.968 217	β^-	2.696 d
80	Mercury	199	^{199}Hg	198.968 253	16.87%	
		202	^{202}Hg	201.970 617	29.86%	
81	Thallium	205	^{205}Tl	204.974 401	70.48%	
82	Lead	206	^{206}Pb	205.974 440	24.1%	
		207	^{207}Pb	206.975 872	22.1%	
		208	^{208}Pb	207.976 627	52.4%	
		210	^{210}Pb	209.984 163	α, β^-	22.3 y
		211	^{211}Pb	210.988 735	β^-	36.1 min
		212	^{212}Pb	211.991 871	β^-	10.64 h
83	Bismuth	209	^{209}Bi	208.980 374	100%	
		211	^{211}Bi	210.987 255	α, β^-	2.14 min
84	Polonium	210	^{210}Po	209.982 848	α	138.38 d
85	Astatine	218	^{218}At	218.008 684	α, β^-	1.6 s
86	Radon	222	^{222}Rn	222.017 570	α	3.82 d
87	Francium	223	^{223}Fr	223.019 733	α, β^-	21.8 min
88	Radium	226	^{226}Ra	226.025 402	α	$1.60 \times 10^3 \text{ y}$

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
89	Actinium	227	^{227}Ac	227.027 750	α, β^-	21.8 y
90	Thorium	228	^{228}Th	228.028 715	α	1.91 y
		232	^{232}Th	232.038 054	100%, α	$1.41 \times 10^{10}\text{y}$
91	Protactinium	231	^{231}Pa	231.035 880	α	$3.28 \times 10^4\text{y}$
92	Uranium	233	^{233}U	233.039 628	α	$1.59 \times 10^3\text{y}$
		235	^{235}U	235.043 924	0.720%, α	$7.04 \times 10^8\text{y}$
		236	^{236}U	236.045 562	α	$2.34 \times 10^7\text{y}$
		238	^{238}U	238.050 784	99.2745%, α	$4.47 \times 10^9\text{y}$
		239	^{239}U	239.054 289	β^-	23.5 min
93	Neptunium	239	^{239}Np	239.052 933	β^-	2.355 d
94	Plutonium	239	^{239}Pu	239.052 157	α	$2.41 \times 10^4\text{y}$
95	Americium	243	^{243}Am	243.061 375	α , fission	$7.37 \times 10^3\text{y}$
96	Curium	245	^{245}Cm	245.065 483	α	$8.50 \times 10^3\text{y}$
97	Berkelium	247	^{247}Bk	247.070 300	α	$1.38 \times 10^3\text{y}$
98	Californium	249	^{249}Cf	249.074 844	α	351 y
99	Einsteinium	254	^{254}Es	254.088 019	α, β^-	276 d
100	Fermium	253	^{253}Fm	253.085 173	EC, α	3.00 d

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
101	Mendelevium	255	^{255}Md	255.091 081	EC, α	27 min
102	Nobelium	255	^{255}No	255.093 260	EC, α	3.1 min
103	Lawrencium	257	^{257}Lr	257.099 480	EC, α	0.646 s
104	Rutherfordium	261	^{261}Rf	261.108 690	α	1.08 min
105	Dubnium	262	^{262}Db	262.113 760	α , fission	34 s
106	Seaborgium	263	^{263}Sg	263.11 86	α , fission	0.8 s
107	Bohrium	262	^{262}Bh	262.123 1	α	0.102 s
108	Hassium	264	^{264}Hs	264.128 5	α	0.08 ms
109	Meitnerium	266	^{266}Mt	266.137 8	α	3.4 ms

Atomic Masses

Selected Radioactive Isotopes

Decay modes are α , β^- , β^+ , electron capture (EC) and isomeric transition (IT). EC results in the same daughter nucleus as would β^+ decay. IT is a transition from a metastable excited state. Energies for β^\pm decays are the maxima; average energies are roughly one-half the maxima.

Isotope	$t_{1/2}$	DecayMode(s)	Energy(MeV)	Percent		γ -Ray Energy(MeV)
^3H	12.33 y	β^-	0.0186	100%		
^{14}C	5730 y	β^-	0.156	100%		
^{13}N	9.96 min	β^+	1.20	100%		
^{22}Na	2.602 y	β^+	0.55	90%	γ	1.27
^{32}P	14.28 d	β^-	1.71	100%		
^{35}S	87.4 d	β^-	0.167	100%		
^{36}Cl	$3.00 \times 10^5 \text{ y}$	β^-	0.710	100%		
^{40}K	$1.28 \times 10^9 \text{ y}$	β^-	1.31	89%		
^{43}K	22.3 h	β^-	0.827	87%	$\gamma \text{ s}$	0.373
						0.618
^{45}Ca	165 d	β^-	0.257	100%		
^{51}Cr	27.70 d	EC			γ	0.320
^{52}Mn	5.59d	β^+	3.69	28%	$\gamma \text{ s}$	1.33
						1.43
^{52}Fe	8.27 h	β^+	1.80	43%		0.169
						0.378
^{59}Fe	44.6 d	$\beta^- \text{ s}$	0.273	45%	$\gamma \text{ s}$	1.10
			0.466	55%		1.29
^{60}Co	5.271 y	β^-	0.318	100%	$\gamma \text{ s}$	1.17
						1.33
^{65}Zn	244.1 d	EC			γ	1.12

Isotope	$t_{1/2}$	DecayMode(s)	Energy(MeV)	Percent		γ -Ray Energy(MeV)
^{67}Ga	78.3 h	EC			γ s	0.0933
						0.185
						0.300
						others
^{75}Se	118.5 d	EC			γ s	0.121
						0.136
						0.265
						0.280
						others
^{86}Rb	18.8 d	β^- s	0.69	9%	γ	1.08
			1.77	91%		
^{85}Sr	64.8 d	EC			γ	0.514
^{90}Sr	28.8 y	β^-	0.546	100%		
^{90}Y	64.1 h	β^-	2.28	100%		
$^{99\text{m}}\text{Tc}$	6.02 h	IT			γ	0.142
$^{113\text{m}}\text{In}$	99.5 min	IT			γ	0.392
^{123}I	13.0 h	EC			γ	0.159
^{131}I	8.040 d	β^- s	0.248	7%	γ s	0.364
			0.607	93%		others
			others			
^{129}Cs	32.3 h	EC			γ s	0.0400
						0.372
						0.411
						others
^{137}Cs	30.17 y	β^- s	0.511	95%	γ	0.662
			1.17	5%		

Isotope	$t_{1/2}$	DecayMode(s)	Energy(MeV)	Percent		γ -Ray Energy(MeV)
^{140}Ba	12.79 d	β^-	1.035	$\approx 100\%$	γ s	0.030
						0.044
						0.537
						others
^{198}Au	2.696 d	β^-	1.161	$\approx 100\%$	γ	0.412
^{197}Hg	64.1 h	EC			γ	0.0733
^{210}Po	138.38 d	α	5.41	100%		
^{226}Ra	$1.60 \times 10^3 \text{ y}$	α s	4.68	5%	γ	0.186
			4.87	95%		
^{235}U	$7.038 \times 10^8 \text{ y}$	α	4.68	$\approx 100\%$	γ s	numerous
^{238}U	$4.468 \times 10^9 \text{ y}$	α s	4.22	23%	γ	0.050
			4.27	77%		
^{237}Np	$2.14 \times 10^6 \text{ y}$	α s	numerous		γ s	numerous
			4.96 (max.)			
^{239}Pu	$2.41 \times 10^4 \text{ y}$	α s	5.19	11%	γ s	7.5×10^{-5}
			5.23	15%		0.013
			5.24	73%		0.052
						others
^{243}Am	$7.37 \times 10^3 \text{ y}$	α s	Max. 5.44		γ s	0.075
			5.37	88%		others
			5.32	11%		
			others			

Selected Radioactive Isotopes

Useful Information

This appendix is broken into several tables.

- [\[link\]](#), Important Constants
- [\[link\]](#), Submicroscopic Masses
- [\[link\]](#), Solar System Data
- [\[link\]](#), Metric Prefixes for Powers of Ten and Their Symbols
- [\[link\]](#), The Greek Alphabet
- [\[link\]](#), SI units
- [\[link\]](#), Selected British Units
- [\[link\]](#), Other Units
- [\[link\]](#), Useful Formulae

Symbol	Meaning	Best Value	Approximate Value
c	Speed of light in vacuum	$2.99792458 \times 10^8 \text{ m/s}$	$3.00 \times 10^8 \text{ m/s}$
G	Gravitational constant	$6.67408(31) \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$	$6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$
N_A	Avogadro's number	$6.02214129(27) \times 10^{23}$	6.02×10^{23}
k	Boltzmann's constant	$1.3806488(13) \times 10^{-23} \text{ J/K}$	$1.38 \times 10^{-23} \text{ J/K}$
R	Gas constant	$8.3144621(75) \text{ J/mol} \cdot \text{K}$	$8.31 \text{ J/mol} \cdot \text{K} = 1.99 \text{ cal/mol} \cdot \text{K} =$
σ	Stefan-Boltzmann constant	$5.670373(21) \times 10^{-8} \text{ W/m}^2 \cdot \text{K}$	$5.67 \times 10^{-8} \text{ W/m}^2 \cdot \text{K}$
k	Coulomb force constant	$8.987551788... \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$	$8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$
q_e	Charge on electron	$-1.602176565(35) \times 10^{-19} \text{ C}$	$-1.60 \times 10^{-19} \text{ C}$
ϵ_0	Permittivity of free space	$8.854187817... \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$	$8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$
μ_0	Permeability of free space	$4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$	$1.26 \times 10^{-6} \text{ T} \cdot \text{m/A}$
h	Planck's constant	$6.62606957(29) \times 10^{-34} \text{ J} \cdot \text{s}$	$6.63 \times 10^{-34} \text{ J} \cdot \text{s}$

Important Constants^[footnote]

Stated values are according to the National Institute of Standards and Technology Reference on Constants, Units, and Uncertainty, www.physics.nist.gov/cuu (accessed May 18, 2012). Values in parentheses are the uncertainties in the last digits. Numbers without uncertainties are exact as defined.

Symbol	Meaning	Best Value	Approximate Value
m_e	Electron mass	$9.10938291(40) \times 10^{-31}\text{kg}$	$9.11 \times 10^{-31}\text{kg}$
m_p	Proton mass	$1.672621777(74) \times 10^{-27}\text{kg}$	$1.6726 \times 10^{-27}\text{kg}$
m_n	Neutron mass	$1.674927351(74) \times 10^{-27}\text{kg}$	$1.6749 \times 10^{-27}\text{kg}$
u	Atomic mass unit	$1.660538921(73) \times 10^{-27}\text{kg}$	$1.6605 \times 10^{-27}\text{kg}$

Submicroscopic Masses^[footnote]

Stated values are according to the National Institute of Standards and Technology Reference on Constants, Units, and Uncertainty, www.physics.nist.gov/cuu (accessed May 18, 2012). Values in parentheses are the uncertainties in the last digits. Numbers without uncertainties are exact as defined.

Sun	mass	$1.99 \times 10^{30}\text{kg}$
	average radius	$6.96 \times 10^8\text{m}$
	Earth-sun distance (average)	$1.496 \times 10^{11}\text{m}$
Earth	mass	$5.9736 \times 10^{24}\text{kg}$
	average radius	$6.376 \times 10^6\text{m}$
	orbital period	$3.16 \times 10^7\text{s}$

Epsilon	Ε	ε	Lambda	Λ	λ	Rho	Ρ	ρ	Psi	Ψ	ψ
Zeta	Ζ	ζ	Mu	Μ	μ	Sigma	Σ	σ	Omega	Ω	ω

The Greek Alphabet

	Entity	Abbreviation	Name
Fundamental units	Length	m	meter
	Mass	kg	kilogram
	Time	s	second
	Current	A	ampere
Supplementary unit	Angle	rad	radian
Derived units	Force	$N = kg \cdot m/s^2$	newton
	Energy	$J = kg \cdot m^2/s^2$	joule
	Power	$W = J/s$	watt
	Pressure	$Pa = N/m^2$	pascal
	Frequency	$Hz = 1/s$	hertz
	Electronic potential	$V = J/C$	volt
	Capacitance	$F = C/V$	farad
	Charge	$C = s \cdot A$	coulomb
	Resistance	$\Omega = V/A$	ohm

	Entity	Abbreviation	Name
	Magnetic field	$T = N/(A \cdot m)$	tesla
	Nuclear decay rate	$Bq = 1/s$	becquerel

SI Units

Length	1 inch (in.) = 2.54 cm (exactly)
	1 foot (ft) = 0.3048 m
	1 mile (mi) = 1.609 km
Force	1 pound (lb) = 4.448 N
Energy	1 British thermal unit (Btu) = 1.055×10^3 J
Power	1 horsepower (hp) = 746 W
Pressure	$1 \text{ lb/in}^2 = 6.895 \times 10^3 \text{ Pa}$

Selected British Units

Length	1 light year (ly) = 9.46×10^{15} m
	1 astronomical unit (au) = 1.50×10^{11} m
	1 nautical mile = 1.852 km
	1 angstrom(\AA) = 10^{-10} m
Area	1 acre (ac) = 4.05×10^3 m ²
	1 square foot (ft ²) = 9.29×10^{-2} m ²
	1 barn (<i>b</i>) = 10^{-28} m ²
Volume	1 liter (<i>L</i>) = 10^{-3} m ³

	1 U.S. gallon (gal) = $3.785 \times 10^{-3} \text{ m}^3$
Mass	1 solar mass = $1.99 \times 10^{30} \text{ kg}$
	1 metric ton = 10^3 kg
	1 atomic mass unit (u) = $1.6605 \times 10^{-27} \text{ kg}$
Time	1 year (y) = $3.16 \times 10^7 \text{ s}$
	1 day (d) = 86,400 s
Speed	1 mile per hour (mph) = 1.609 km/h
	1 nautical mile per hour (naut) = 1.852 km/h
Angle	1 degree ($^\circ$) = $1.745 \times 10^{-2} \text{ rad}$
	1 minute of arc ($'$) = 1/60 degree
	1 second of arc ($''$) = 1/60 minute of arc
	1 grad = $1.571 \times 10^{-2} \text{ rad}$
Energy	1 kiloton TNT (kT) = $4.2 \times 10^{12} \text{ J}$
	1 kilowatt hour ($\text{kW} \cdot \text{h}$) = $3.60 \times 10^6 \text{ J}$
	1 food calorie (kcal) = 4186 J
	1 calorie (cal) = 4.186 J
	1 electron volt (eV) = $1.60 \times 10^{-19} \text{ J}$
Pressure	1 atmosphere (atm) = $1.013 \times 10^5 \text{ Pa}$
	1 millimeter of mercury (mm Hg) = 133.3 Pa
	1 torricelli (torr) = 1 mm Hg = 133.3 Pa
Nuclear decay rate	1 curie (Ci) = $3.70 \times 10^{10} \text{ Bq}$

Other Units

Circumference of a circle with radius r or diameter d	$C = 2\pi r = \pi d$
Area of a circle with radius r or diameter d	$A = \pi r^2 = \pi d^2/4$
Area of a sphere with radius r	$A = 4\pi r^2$

Volume of a sphere with radius r

$$V = \frac{4}{3} \pi r^3$$

Useful Formulae

Glossary of Key Symbols and Notation

In this glossary, key symbols and notation are briefly defined.

Symbol	Definition
any symbol	average (indicated by a bar over a symbol— e.g., \bar{v} is average velocity)
$^{\circ}\text{C}$	Celsius degree
$^{\circ}\text{F}$	Fahrenheit degree
//	parallel
\perp	perpendicular
\propto	proportional to
\pm	plus or minus

Symbol	Definition
0	zero as a subscript denotes an initial value
α	alpha rays
α	angular acceleration
α	temperature coefficient(s) of resistivity
β	beta rays
β	sound level
β	volume coefficient of expansion
β^{-}	electron emitted in nuclear beta decay
β^{+}	positron decay
γ	gamma rays

Symbol	Definition
γ	surface tension
$\gamma = 1/\sqrt{1 - v^2/c^2}$	a constant used in relativity
Δ	change in whatever quantity follows
δ	uncertainty in whatever quantity follows
ΔE	change in energy between the initial and final orbits of an electron in an atom
ΔE	uncertainty in energy
Δm	difference in mass between initial and final products
ΔN	number of decays that occur
Δp	change in momentum

Symbol	Definition
Δp	uncertainty in momentum
ΔPE_g	change in gravitational potential energy
$\Delta \theta$	rotation angle
Δs	distance traveled along a circular path
Δt	uncertainty in time
Δt_0	proper time as measured by an observer at rest relative to the process
ΔV	potential difference
Δx	uncertainty in position
ϵ_0	permittivity of free space
η	viscosity

Symbol	Definition
θ	angle between the force vector and the displacement vector
θ	angle between two lines
θ	contact angle
θ	direction of the resultant
θ_b	Brewster's angle
θ_c	critical angle
κ	dielectric constant
λ	decay constant of a nuclide
λ	wavelength
λ_n	wavelength in a medium

Symbol	Definition
μ_0	permeability of free space
μ_k	coefficient of kinetic friction
μ_s	coefficient of static friction
ν_e	electron neutrino
π^+	positive pion
π^-	negative pion
π^0	neutral pion
ρ	density
ρ_c	critical density, the density needed to just halt universal expansion
ρ_{fl}	fluid density

Symbol	Definition
ρ_{obj}	average density of an object
ρ/ρ_{w}	specific gravity
τ	characteristic time constant for a resistance and inductance (RL) or resistance and capacitance (RC) circuit
τ	characteristic time for a resistor and capacitor (RC) circuit
τ	torque
Υ	upsilon meson
Φ	magnetic flux
ϕ	phase angle
Ω	ohm (unit)
ω	angular velocity

Symbol	Definition
A	ampere (current unit)
A	area
A	cross-sectional area
A	total number of nucleons
a	acceleration
a_B	Bohr radius
a_c	centripetal acceleration
a_t	tangential acceleration
AC	alternating current
AM	amplitude modulation

Symbol	Definition
atm	atmosphere
B	baryon number
B	blue quark color
B	antiblue (yellow) antiquark color
b	quark flavor bottom or beauty
B	bulk modulus
B	magnetic field strength
B_{int}	electron's intrinsic magnetic field
B_{orb}	orbital magnetic field
BE	binding energy of a nucleus—it is the energy required to completely disassemble it into separate protons and neutrons

Symbol	Definition
BE/A	binding energy per nucleon
Bq	becquerel—one decay per second
C	capacitance (amount of charge stored per volt)
C	coulomb (a fundamental SI unit of charge)
C_p	total capacitance in parallel
C_s	total capacitance in series
CG	center of gravity
CM	center of mass
c	quark flavor charm
c	specific heat

Symbol	Definition
c	speed of light
Cal	kilocalorie
cal	calorie
COP_{hp}	heat pump's coefficient of performance
COP_{ref}	coefficient of performance for refrigerators and air conditioners
$\cos \theta$	cosine
$\cot \theta$	cotangent
$\csc \theta$	cosecant
D	diffusion constant
d	displacement

Symbol	Definition
d	quark flavor down
dB	decibel
d_i	distance of an image from the center of a lens
d_o	distance of an object from the center of a lens
DC	direct current
E	electric field strength
ε	emf (voltage) or Hall electromotive force
emf	electromotive force
E	energy of a single photon
E	nuclear reaction energy

Symbol	Definition
E	relativistic total energy
E	total energy
E_0	ground state energy for hydrogen
E_0	rest energy
EC	electron capture
E_{cap}	energy stored in a capacitor
Eff	efficiency—the useful work output divided by the energy input
Eff _C	Carnot efficiency
E_{in}	energy consumed (food digested in humans)
E_{ind}	energy stored in an inductor

Symbol	Definition
E_{out}	energy output
e	emissivity of an object
e^+	antielectron or positron
eV	electron volt
F	farad (unit of capacitance, a coulomb per volt)
F	focal point of a lens
F	force
F	magnitude of a force
F	restoring force
F_{B}	buoyant force

Symbol	Definition
F_c	centripetal force
F_i	force input
\mathbf{F}_{net}	net force
F_o	force output
FM	frequency modulation
f	focal length
f	frequency
f_0	resonant frequency of a resistance, inductance, and capacitance (RLC) series circuit
f_0	threshold frequency for a particular material (photoelectric effect)

Symbol	Definition
f_1	fundamental
f_2	first overtone
f_3	second overtone
f_B	beat frequency
f_k	magnitude of kinetic friction
f_s	magnitude of static friction
G	gravitational constant
G	green quark color
\bar{G}	antigreen (magenta) antiquark color

Symbol	Definition
g	acceleration due to gravity
g	gluons (carrier particles for strong nuclear force)
h	change in vertical position
h	height above some reference point
h	maximum height of a projectile
h	Planck's constant
hf	photon energy
h_i	height of the image
h_o	height of the object
I	electric current

Symbol	Definition
I	intensity
I	intensity of a transmitted wave
I	moment of inertia (also called rotational inertia)
I_0	intensity of a polarized wave before passing through a filter
I_{ave}	average intensity for a continuous sinusoidal electromagnetic wave
I_{rms}	average current
J	joule
J/Ψ	Joules/psi meson
K	kelvin
k	Boltzmann constant

Symbol	Definition
k	force constant of a spring
K_{α}	x rays created when an electron falls into an $n = 1$ shell vacancy from the $n = 3$ shell
K_{β}	x rays created when an electron falls into an $n = 2$ shell vacancy from the $n = 3$ shell
kcal	kilocalorie
KE	translational kinetic energy
KE + PE	mechanical energy
KE_e	kinetic energy of an ejected electron
KE_{rel}	relativistic kinetic energy
KE_{rot}	rotational kinetic energy
KE	thermal energy

Symbol	Definition
kg	kilogram (a fundamental SI unit of mass)
L	angular momentum
L	liter
L	magnitude of angular momentum
L	self-inductance
ℓ	angular momentum quantum number
L_{α}	x rays created when an electron falls into an $n = 2$ shell from the $n = 3$ shell
L_e	electron total family number
L_{μ}	muon family total number
L_{τ}	tau family total number

Symbol	Definition
L_f	heat of fusion
L_f and L_v	latent heat coefficients
L_{orb}	orbital angular momentum
L_s	heat of sublimation
L_v	heat of vaporization
L_z	z - component of the angular momentum
M	angular magnification
M	mutual inductance
m	indicates metastable state
m	magnification

Symbol	Definition
m	mass
m	mass of an object as measured by a person at rest relative to the object
m	meter (a fundamental SI unit of length)
m	order of interference
m	overall magnification (product of the individual magnifications)
$m(^AX)$	atomic mass of a nuclide
MA	mechanical advantage
m_e	magnification of the eyepiece
m_e	mass of the electron
m_ℓ	angular momentum projection quantum number

Symbol	Definition
m_n	mass of a neutron
m_o	magnification of the objective lens
mol	mole
m_p	mass of a proton
m_s	spin projection quantum number
N	magnitude of the normal force
N	newton
N	normal force
N	number of neutrons
n	index of refraction

Symbol	Definition
n	number of free charges per unit volume
N_A	Avogadro's number
N_r	Reynolds number
$N \cdot m$	newton-meter (work-energy unit)
$N \cdot m$	newtons times meters (SI unit of torque)
OE	other energy
P	power
P	power of a lens
P	pressure
p	momentum

Symbol	Definition
p	momentum magnitude
p	relativistic momentum
\mathbf{p}_{tot}	total momentum
\mathbf{p}'_{tot}	total momentum some time later
P_{abs}	absolute pressure
P_{atm}	atmospheric pressure
P_{atm}	standard atmospheric pressure
PE	potential energy
PE _{el}	elastic potential energy
PE _{elec}	electric potential energy

Symbol	Definition
PE_s	potential energy of a spring
P_g	gauge pressure
P_{in}	power consumption or input
P_{out}	useful power output going into useful work or a desired, form of energy
Q	latent heat
Q	net heat transferred into a system
Q	flow rate—volume per unit time flowing past a point
$+Q$	positive charge
$-Q$	negative charge

Symbol	Definition
q	electron charge
q_p	charge of a proton
q	test charge
QF	quality factor
R	activity, the rate of decay
R	radius of curvature of a spherical mirror
R	red quark color
R	antired (cyan) quark color
R	resistance
R	resultant or total displacement

Symbol	Definition
R	Rydberg constant
R	universal gas constant
r	distance from pivot point to the point where a force is applied
r	internal resistance
r_{\perp}	perpendicular lever arm
r	radius of a nucleus
r	radius of curvature
r	resistivity
r or rad	radiation dose unit
rem	roentgen equivalent man

Symbol	Definition
rad	radian
RBE	relative biological effectiveness
RC	resistor and capacitor circuit
rms	root mean square
r_n	radius of the n th H-atom orbit
R_p	total resistance of a parallel connection
R_s	total resistance of a series connection
R_s	Schwarzschild radius
S	entropy
S	intrinsic spin (intrinsic angular momentum)

Symbol	Definition
S	magnitude of the intrinsic (internal) spin angular momentum
S	shear modulus
S	strangeness quantum number
s	quark flavor strange
s	second (fundamental SI unit of time)
s	spin quantum number
s	total displacement
$\sec \theta$	secant
$\sin \theta$	sine
s_z	z-component of spin angular momentum

Symbol	Definition
T	period—time to complete one oscillation
T	temperature
T_c	critical temperature—temperature below which a material becomes a superconductor
T	tension
T	tesla (magnetic field strength B)
t	quark flavor top or truth
t	time
$t_{1/2}$	half-life—the time in which half of the original nuclei decay
$\tan \theta$	tangent
U	internal energy

Symbol	Definition
u	quark flavor up
u	unified atomic mass unit
u	velocity of an object relative to an observer
u'	velocity relative to another observer
V	electric potential
V	terminal voltage
V	volt (unit)
V	volume
v	relative velocity between two observers
v	speed of light in a material

Symbol	Definition
\mathbf{v}	velocity
\mathbf{v}	average fluid velocity
$V_B - V_A$	change in potential
\mathbf{v}_d	drift velocity
V_p	transformer input voltage
V_{rms}	rms voltage
V_s	transformer output voltage
\mathbf{v}_{tot}	total velocity
v_w	propagation speed of sound or other wave
\mathbf{v}_w	wave velocity

Symbol	Definition
W	work
W	net work done by a system
W	watt
w	weight
w_{fl}	weight of the fluid displaced by an object
W_{c}	total work done by all conservative forces
W_{nc}	total work done by all nonconservative forces
W_{out}	useful work output
X	amplitude
X	symbol for an element

Symbol	Definition
${}_A^ZX_N$	notation for a particular nuclide
x	deformation or displacement from equilibrium
x	displacement of a spring from its undeformed position
x	horizontal axis
X_C	capacitive reactance
X_L	inductive reactance
x_{rms}	root mean square diffusion distance
y	vertical axis
Y	elastic modulus or Young's modulus
Z	atomic number (number of protons in a nucleus)

Symbol	Definition
Z	impedance